U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 191

How to Pick a New Forecast Model:
The Selection of the 7L PE for NMC Operations

John D. Stackpole, Armand Desmarais, John Hovermale,
Robert van Haaren, Norman Phillips, Ken Campana,
Robert Hopkins, and Staff Members, NMC

OCTOBER 1978

I. Introduction

During the better part of 1977 the Development Division of the National Meteorological Center undertook a series of intercomparison tests of three hemispheric numerical weather prediction models. The aim was to select one as a replacement for the then operational 6-layer primitive equation model (6LPE) (Shuman and Hovermale, 1968) in use, with modifications, since June of 1966. What follows is short descriptions of the three contending models, a description of the nature of the tests, an account of their execution, and the results.

In addition to the obvious purpose of documenting the reasons for a change in NMC's operational models, this essay may serve also as an example of the nature and dimensions of such tests. Tests like this one, whether they be forecast model, analysis system, data system, or operational configuration intercomparisons are of considerable magnitude and are not to be undertaken for light and transient reasons. What follows may give some indication of the amount of effort required to answer the deceptively simple question: "Is this model better than that one?" It may also serve as a guide of how to answer similar questions in the future.

II. The Three Model Runoff Tests

A. The Contenders

Three contenders for replacement of the 6LPE were included in the initial series of tests. In no particular order they were:

1. The Nested Grid Model (NGM).

The salient features of this model (under continuing development by N. A. Phillips) at the time of the tests were: 9 sigma layers from the model terrain surface to 0 mb; a hemispheric coarse mesh grid on a polar stereographic projection true at 60°N (grid size 410 km at 45°N) in which is embedded a rectangular fine mesh sub-area of one-half the grid size, centered over North America and adjacent waters; and much of the "usual" physics: orography, surface drag, Kuo-type convective precipitation and large scale condensation.

2. Nine layer 2° λ-ϕ Hemispheric (9LH)

This is a variant of the 9 layer global model (Stackpole, 1978) used in NMC operations. The alterations consist in changing the grid spacing from 2.5° to 2° and restricting the horizontal domain to the northern hemisphere. Otherwise the model characteristics are the same: 9 sigma layers (6 tropospheric, 3 stratospheric, with a material surface "tropopause" between them) from the terrain to 50 mb; 2° lola (longitude-latitude) spherical coordinate grid; and much the same physics: orography, surface friction, large scale and convective precipitation, radiative heating and cooling and various surface boundary effects.

3. Hemispheric Fine Mesh (HFM)

This is an exact duplicate of the operational 6LPE model except it uses a grid with one half the mesh length - 190.5 km, at 60^{o} north, rather than 381 km. As a reminder to the reader the HFM has a boundary layer, three tropospheric and two stratospheric layers separated by a material surface "tropopause"; the top varies around 70-80 mb; the grid is a 129x129 point mesh on a polar stereographic map; and the additional physics includes orography, surface friction, large scale precipitation, and highly simplified convective rain and radiative effects.

A major constraint on the various design decisions in these models was that of computer running time: For a model to be eligible to replace the 6LPE it had to require not more (or not much more) than three times the 6LPE run time - no more than some 27 minutes computer time for a 24-hour forecast.

The NGM and 9LH were both constructed to fit the 27 min/24 hour constraint, the inclusion of the HFM in the runoff was made possible by the development of a numerical device, "pressure gradient averaging", (Brown and Campana, 1978) which made it possible to cut the grid size without reducing the time step.

During the various tests the 6LPE was itself a (fourth) contender in the runoff - there was no *à priori* assumption that any of the new models could best the 6LPE. There was some fair degree of hope, of course, making it worthwhile to undertake the tests.

B. Design of Test

1. Operational Uses of 6LPE Model

In that the purpose of the test was to select a possible replacement for the currently operational 6-layer PE model, it was the first order of business to decide to what specific purposes the 6LPE model is currently put. The LFM model serves as the principal (but not exclusive) source of forecast guidance for the contiguous U.S. for up to 48 hours, leaving guidance for other areas of the hemisphere and further out in time to the 6LPE model. Also there are some aspects of the contiguous U.S. forecasts from the LFM model that leave something to be desired (e.g., precipitation forecasts, the "locked in" error) and any improvement that could be obtained from a replacement model in these respects would be most welcome. Such considerations as these plus general forecaster experience with the 6LPE model led us to pay particular attention to:

- The sea level pressure and 500-1000 mb thickness forecasts for 24 and 48 hours over the Western Atlantic area, and an area loosely specified as the Eastern Pacific, Alaska, and Western U.S. These forecasts are in support of NMC's marine responsibilities and also relate to the feeling, expressed by some users, that the LFM forecasts are sometimes deficient over the Western U.S. and Alaska.

- The jet stream at 300 mb over the Pacific, U.S., and Atlantic. This, of course, relates to NMC's aviation forecast responsibilities.

- The quality, both in terms of coverage and amounts, of the precipitation forecasts over the Eastern and Western U.S. (divided at longitude 105°W) at 24 and 48 hours.

- The overall quality (500 mb height, sea level pressure and 500-1000 mb thickness, precipitation as in the previous category) of the 84-hour forecast--this supports the NMC extended forecast program.

- The general quality, (at 24, 48, and 84 hours) consistency, and credibility of the forecasts, with particular reference to the 500 mb heights and vorticity. This relates to NMC's responsibility to issue forecasts that make sense and can be useful to the field.

- The area of the 6LPE forecast domain in the vicinity of the LFM boundaries. Any replacement for the 6LPE will have to supply the time varying boundary conditions for the LFM model, and thus should make an adequate forecast of them.

- The wind and temperature forecasts at 24 hours for 100 mb in the area of the SuperSonic Transport flight paths. At present persistence rather than the 6LPE forecast is employed for the "forecast"; a model that can do better than persistence at this level will be most welcome.

In addition to this list of operational uses, some attention was given to how well the potential replacement models could correct certain characteristic errors of the 6LPE model, in particular:

- The "locked-in error" in which the model fails to bring a low and trough out of the south or southwest portion of the U.S.;

- "Cross contour flow" in which forecast isotach maxima cross height contours at a large angle; and

- Inconsistent patterns of height and vorticity contours.

In a sense, these three problem areas are special cases of the "general quality, consistency, and credibility" category outlined above.

2. Selection of Cases

In an ideal world the selection of test cases would involve simply turning to a library of historical situations and picking out those that exemplified the particular uses of the forecasts outlined above. Unfortunately, this was not possible. The library situations were generally saved on the basis of continental U.S. meteorological phenomena: Big storms and other special situations or else (sometimes "and") because one of the operational models (LFM or 6LPE) showed a particularly bad error. The cases then were selected without much attention to their extra-U.S. and post-48 hour characteristics but as many as possible of the various usage criteria that could be applied were applied in the selection process. The unstated assumption was that what was good for the U.S. would be good for the rest of the Northern Hemisphere. One case was selected in part because it was a good 6LPE forecast--it wouldn't do to deteriorate good forecasts while questing after improvements in bad ones. Economics and constraints of time dictated that only 6 or so cases could be run (to 84 hours) for intercomparisons between the three contenders and comparison with the operational 6LPE forecasts; Table II-1 indicates the initial times and some of the salient features of the cases. A couple of explanatory notes: The February 17 case was selected solely because of an extreme example of "cross contour flow" in the Pacific and was not included in all of the general evaluations; the heading "Western U.S. Problems" generally means poor forecasts in that area and relates directly to the 6LPE usage criteria dealing with Marine and Western U.S. forecast responsibilities. The cases were selected from the available set by a committee of H. Saylor, N. Phillips, J. Hovermale, J. Stackpole, J. Brown, R. Hopkins, and S. Tracton.

CASE		CROSS-CONTOUR FLOW	LOCKED-IN ERROR	CYCLOGENESIS	PRECIPITATION	VORTICITY, HEIGHT OUT OF PHASE	WESTERN U.S. PROBLEMS
April 18, 1975 (00Z) (Bad 6LPE)	1		✓		✓		
August 23, 1975 (00Z)	2			✓	✓		
Nov. 24, 1976 (00Z) (Good 6LPE, Bad LFM)	3			✓			✓
Dec. 8, 1976 (12Z)	4		✓	✓	✓		✓
Jan 9, 1977 (00Z)	5	✓	✓	✓		✓	
Feb. 21, 1977 (12Z)	6			✓	✓	✓	✓
Feb. 17, 1977 (00Z) (36-hr fcst)	7	✓					

Table II-1
Selected Cases and Salient Features

All the forecasts were run to 84 hours from the same Hough analysis initial conditions with maps and other output readied at 24, 48 and 84 hours. Two of the cases, 18 Apr and 23 Aug, were extended to 5 days with maps made at 120 hours. These latter were to give at least a preliminary indication of the possible value of the new models in the extended range.

3. Subjective Evaluation

For the subjective evaluations an experienced jury of NMC Forecast Division forecasters (H. Saylor, J. O'Connor, H. Brown, D. Olson, and R. McCarter) were given maps of:

- Sea level pressure and 500-1000 mb thickness
- 500 mb heights and vorticity
- 300 mb heights and isotachs
- 500 mb height error (forecast minus analysis)
- 12 hour accumulated precipitation
- The verifying analyses

One set of these maps were made for each forecast hour (24, 48, and 84) and for each of the three contending models plus the operational 6LPE model. On the maps only the 6LPE model was identified as such--the other sets of forecasts bore only a coded indicator (A, B or C) of the forecast model. For any particular case the code indicated the same model for the three forecast hours but the code was changed, randomly, from case to case. Thus the forecaster jury was discouraged from prejudging any particular model.

The maps were the usual "Varian" maps of the style and size routinely used at NMC with one exception, the precipitation forecasts. Since Varian mapping codes work with the 381 km $1:30 \times 10^6$ scale polar stereographic map projection grid, the final output from each of the contending models had to be interpolated from the individual model's forecasting grid to that grid. The precipitation forecasts, being discontinuous, would suffer considerable debasement in the interpolation process and a different procedure was followed: The models' precipitation forecasts were printed on the scale of the particular model grid, transcribed to a suitable map base, hand analyzed and the analysis traced to another map base. This last map thus contained all of the resolution that the models were capable of, did not show the source of the forecast, and was the map (suitably coded) that went to the forecaster jury.

Along with the maps the jury received a set of instructions and a questionnaire designed to direct their attention toward the operational usage criteria outlined in Section 1 above. They are in Appendix I and, as can be seen, ask the jury members to rank the three contenders and the 6LPE on a number of categories relating to the usage criteria.

An error crept into the form and was not noticed until a number of the jurors had completed their evaluations. The principal NMC responsibility for aviation wind forecasting is for 24 hour forecasts - by error the form requests the wind evaluation at 48 hours. When the error was noticed, one of the jurors (McCarter, the aviation forecaster), was asked to check if going back and correcting the error would make any difference. He reported that it would not - rankings based on the 24 hour forecast would have been substantially the same as the ones done with the 48 hour forecasts. We did not bother, therefore, to ask the jury to reevaluate the 300 mb wind forecasts for 24 hours.

An additional set of maps of 100 mb winds and temperatures was prepared and passed on to the NMC Development Division's Upper Air Branch for their subjective evaluation both for general quality and for particular consideration of the forecasts for the SuperSonic Transport Aircraft routes. No questionnaire was prepared.

4. Objective (statistical) Evaluations

For the objective verifications a more or less standard set of statistics was calculated, to wit: Mean (bias) and root mean square (rms) errors of geopotential heights, temperatures, relative humidity and wind speeds, plus rms vector wind error, Tewles/Wobus S1 Score, and threat score and bias of precipitation forecasts. The forecasts were verified for 24, 48 and 84 hours (though, see below, not all the cases were verified for all the time periods because of missing verification data) and at the 1000 mb, 850 mb, 500 mb, 300 mb, and 100 mb mandatory pressure levels.

The data against which the forecasts were tested were of two kinds: Gridded analyses (The NMC FINAL analysis) and radiosonde upper air observations. (Raob measurements at 1000 mb were not used, thus avoiding problems introduced by various "reduction-to-sea-level" methods when the 1000 mb surface was underground). Anticipating the results somewhat, it became apparent that the verifications-against-analyses and the verifications-against-observations lead to the same conclusions; in the interest of reducing bulk somewhat, the statistics incorporated in this report are those of the verifications against observations only (except at 1000 mb where the analyses were used).

Two networks of observation stations were used: 110 stations over North America (essentially all the regularly reporting Raob stations from 25° to 145° West longitude) and 102 stations quasi-uniformly distributed over the entire Northern Hemisphere. Table II-2 lists the WMO Block and Station number for the stations in the two networks.

110 North American Stations

70361
71811
71907
72225
72261
72340
72393
72476
72576
72694
74486

70398
71815
71913
72229
72265
72349
72402
72486
72583
72712
74494

71109
71816
71934
72232
72270
72353
72403
72493
72597
72734
74794

71115
71826
71945
72235
72274
72355
72425
72518
72606
72747
76151

71119
71836
72201
72240
72290
72363
72429
72520
72637
72764
76225

71399
71848
72203
72247
72291
72365
72433
72528
72645
72768
76255

71600
71853
72208
72250
72304
72374
72451
72532
72654
72775
76394

71701
71867
72210
72255
72311
72381
72456
72553
72655
72785
78016

71722
71896
72213
72257
72317
72385
72468
72562
72662
72797
99200

71801
71906
72220
72260
72327
72391
72469
72572
72681
74207
99215

102 Northern Hemisphere Stations

01001
08521
21982
31088
42647
51709
70133
71925
72469
78016
99217

01028
08536
23472
31329
42809
52418
70326
72203
72493
78954
99227

01384
10338
25677
31960
44292
54511
70361
72208
72532
91066

02935
12982
25954
32540
45004
57127
70414
72232
72606
91165

03005
16242
26702
33837
47412
57679
71072
72250
72747
91245

03322
20046
27037
34009
47646
58367
71600
72290
72775
99195

03953
20292
28661
35108
47827
59758
71815
72304
72797
99200

04018
20744
29612
36177
47971
60680
71836
72340
76151
99211

04202
21824
29865
38062
47991
60715
71909
72363
76458
99212

09509
21965
30230
38750
50527
70086
71913
72403
76644
99215

Table II-2

Verification Station Networks

For the 1000 mb verifications against analyses the (roughly) corresponding North American and hemispheric areas are indicated on Figure II-1 as AREA I and the octagonal outline respectively.

Other networks of observations stations were established and verified - Alaska and vicinity, Western Europe, Eastern Asia, stations along the Washington to London/Paris SST route, stations in the vicinity of the LFM boundaries - but again the conclusions drawn from these verification networks were no different from those of the North American and Northern Hemisphere networks. Only the latter network statistics are included below.

The method of calculation of the various error statistics is straightforward with one exception, the S1 score. For the mean and rms error statistics the forecast quantities were biquadratically interpolated to the station locations, the errors established and the appropriate summations over all the stations with valid reports in the network performed. For the S1 score calculation, a preliminary pass is made through all of the available upper air observations (not just those of the network in question) and the station which is the nearest neighbor to each of the network stations is located. Then the observed and forecast height (and, for the fun of it, temperature) gradients between the station pairs are used for the S1 score calculation. The "nearest neighbor" selection is limited by claustrophobic (pairs closer than 100 km are not allowed), agoraphobic (pairs separated by more than 2357 km are not allowed) and geminiphobic (if A selects B as its closest neighbor, B may not select A) constraints. This method of calculation of S1 differs from the usual one in which the gradients are computed between pre-selected grid points in a fixed geographic array. Again comparison between the station S1 and grid S1 scores for the various forecasts and models showed no significant differences in the conclusions one would draw from them. Excepting 1000 mb, station S1 scores are presented below.

For the objective verification of precipitation forecasts a different network of stations was used - a network of 60 first order stations (long in use by NMC Forecast Division) was augmented by 30 additional stations designed to fill some gaps and cover problem areas along coasts and mountain areas. A computer algorithm was readied, designed to interpolate (in a manner appropriate to the discontinuous precipitation fields) from the grid points at which precipitation was forecast in the various models to these stations. Each of the models incorporated this computation in their output sections, thus producing a list of 12 hour accumulated precipitation amounts for the verification times and stations. These station forecasts were the material for the calculation of the precipitation threat and bias scores.

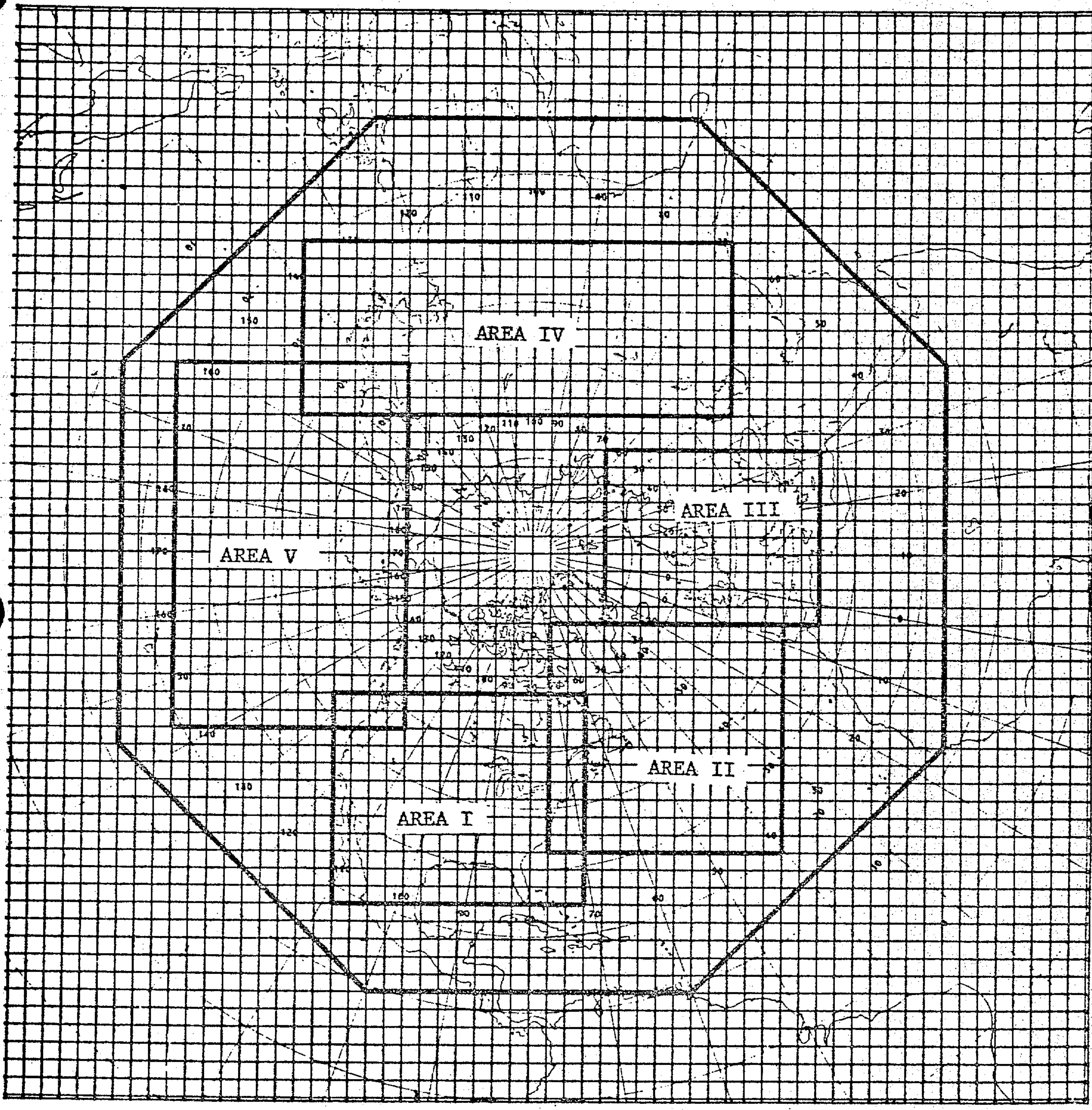


Fig. II-1
Areas for Verifications
against Analyses

C. Execution of the Tests

Each of the researchers associated with the particular contending models (Phillips and Campana for the NGM; Stackpole for the 9LH; Hovermale and Desmarais for the HFM) was responsible for the actual executions of the forecasts with their models and the production of output tapes in a standard form. These tapes were then processed by Desmarais and SEB personnel to produce the objective verification statistics and the multitudes of maps for the subjective verifications.

During the running of the models a few problems (other than unearthing computer programming errors) were encountered, requiring, in some case, quick repairs to the models to enable them to continue the tests.

The 9 layer hemispheric model encountered two such problems. In the first, some preliminary runs revealed that the upper level forecasts 48 hours and beyond exhibited undesirable roughness, particularly in the wind (and vorticity) fields. A diffusive damping filter was hastily constructed and applied to the rotational part of the winds. This "vorticity damper" works on the same principle as the divergence damper proposed by Shuman some years ago (Shuman and Stackpole, 1969). The K constant, which is applied effectively to the laplacian of the vorticity, had the rather large value of 8×10^5 (mks units) throughout the forecasts. The expected palliation was achieved but unfortunately at the cost of excessively reducing wind speed maxima. The second problem for the 9LH arose out of an inconsistency between the initializing of the specific humidity values and some assumptions as to their saturation value in the precipitation forecast sections. When this was fixed, a couple of the forecasts that had shown violent numerical instabilities (particularly in the tropics) induced by excessive latent heat release ran without difficulty.

No changes were made in either the HFM or NGM (or 6LPE, of course) during the course of the tests, although the results, see below, did indicate some modifications to be appropriate.

The original plan called for 6 cases forecast to 84 hours (and a seventh to 36 hours) verified (subjectively and objectively) at 24, 48, and 84 hours with each of the four models. Unfortunately, the vagaries of computer systems (broken magnetic tapes, historical data not saved, etc.) precluded making the verifications at all of the 76 possible opportunities. In summary, subjective verifications were made for all models, cases, and times except the 84 hour 6LPE forecasts from 24 Nov 75, 8 Dec 76 and 21 Feb 77; objective verifications were made at all possible opportunities except all of the 6LPE forecasts from 00Z 9 Jan 1977 and the 84 hour 6LPE forecasts from 00Z 24 Nov 1976, 8 Dec 1976, 00Z 9 Jan 1977 and 12Z 21 Feb 1977; i.e., out of 6 potential 84 hour forecasts two 6LPE forecasts were verified objectively and three subjectively. (The two 12Z initial time forecasts were not operationally run to 84 hours, the 00Z 24 Nov 76 case was not saved through 84 hours at the time, and the tape containing the 00Z 9 Jan 1977 forecast was mutilated by the machine after the maps were made but before the objective verification was done).

As mentioned earlier one of the design criteria for the various computing models was that their run times not be excessive. Table II-3 gives the CPU (Central Processor Unit) and Wall clock times (to the nearest minute) for a 24 hour forecast (without any output map processing) in a computer environment simulating operational conditions.

TIME	9LH	NGM	HFM
CPU	25	28	26
WALL	30	33	32
Ratio	.83	.85	.81

Table II-3 Run times (minutes) for 24 hr forecast.

It is, of course, no surprise that these run times satisfy the operational criteria - had they not, for a particular model, that model would not have been an entry in the evaluation runoff.

All of the very considerable number of maps (some 102 forecast and verification maps for each case) with copies, magnetic tapes, printouts and what all have been saved for possible future reference. The sheer bulk of them suggests, however, that they probably will not be saved indefinitely.

D. Results

The number of cases, verification times, criteria for subjective evaluation, statistics computable (and computed) for objective verification, etc., obviously precludes any detailed presentation and discussion of all of them - instead a selection and abstraction of results will be presented. Jumping ahead a bit, we were fortunate in that the various evaluations were consistent in their conclusions - this made the final decision considerably less onerous than would have been the case otherwise.

1. Objective Statistics

After considerable agonizing over the various statistics computed (the complete set of which are available for inspection at NMC) we settled on an analysis and presentation of five: The S1 height gradient error, the root-mean-square vector wind error, root-mean-square temperature error and precipitation threat score and bias. For the most part the analysis is in terms of the relative ranking of the scores for the four contending models and persistence for the three statistical quantities. If the "forecast" (including persistence) had the best score of the five it was ranked 1, worst 5, etc. These rankings were then established case-by-case, pressure level-by-pressure level, statistic-by-statistic, forecast time-by-forecast time, and then averaged in various combinations.

Table II-4 is the first of these averages- here is shown the numerical rank, averaged for the three tropospheric forecast levels (850, 500 and 300 mb) for each case and each model for the Northern Hemisphere 102 station net. Here it is evident that the preponderance of low numbers (better scores) lies with the HFM; persistence is clearly poorest; the other three models are closely bunched with the 6LPE and NGM perhaps having a slight edge over the 9LH. Table II-5 shows the same set of ranks but for 100 mb only. The situation has quite reversed itself: The HFM has dropped to last place, perhaps sharing that position with the 6LPE, while the 9LH and NGM are vying for first, with persistence close behind. What is also illustrated by Table II-4 and II-5 is that, despite the not inconsiderable case to case variability of the rankings, a clear pattern can be detected - there is a signal in the noise - and thus we can be confident that averages over all the cases will be meaningful. Table II-6 is just such an average of the ranks over all cases but broken out by pressure levels, again for the Northern Hemisphere station network. The conclusions are much the same as before - HFM first, PER last and the other three bunched for the tropospheric levels. At 100 mb the HFM and 6LPE are clearly last, while the 9LH and NGM just nose out persistence. Table II-7 repeats the content of II-6 but for the 110 station North American network and adds the rankings of the 1000 mb S1 score (computed with respect to <u>analyses</u>, not observations) and the ranking of the precipitation scores. Here the observations are altered somewhat. In the troposphere PER is still last, HFM first but now, over North America, the NGM has emerged from the bunch to become a clear and close second. Presumably this can be attributed to the placement of the NGM's fine mesh grid over the North American area - increased resolution is obviously advantageous. In the stratosphere the NGM and 9LH continue to lead, but in contrast to the hemisphere as a whole, the HFM and 6LPE are better than persistence. However, the latter is not true for temperature, which is probably the most important forecast element at 100 mb. The precipitation rankings are something of a disappointment - no clear victor emerges - precipitation forecasting continues to be a tough nut. Table II-8 is a summary and condensation the previous ones - it shows the ranks averaged over time (three forecast hours) space (both areas) and the tropospheric pressure levels. The conclusions, as one should expect, are the same: HFM first, PER last and the other three models clustered. The precipitation shows no particular leader.

Table II-4

CASE-BY-CASE
RANKING OF OBJECTIVE VERIFICATION STATISTICS FOR FOUR HEMISPHERIC MODELS AND PERSISTENCE AVERAGE OF 850-500-300 MB RANKS

NORTHERN HEMISPHERE 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
<u>850-300 MB</u>					
CASE 1	0000 GMT April 18, 1975				
S1	4.0/4.0/3.3	2.0/2.0/2.3	1.3/1.0/1.0	2.3/3.0/3.3	5.0/5.0/5.0
RMSVE	3.0/3.0/2.0	3.3/1.7/2.7	1.0/3.0/1.0	2.7/2.0/4.0	5.0/5.0/5.0
RMSTE	3.7/3.0/3.7	3.0/2.7/3.0	1.0/1.0/1.3	1.7/1.7/1.7	5.0/5.0/5.0
CASE 2	0000 GMT August 23, 1975				
S1	3.0/3.0/3.0	2.3/3.3/2.0	1.3/1.0/1.0	3.0/2.7/4.3	5.0/5.0/4.7
RMSVE	2.7/2.3/1.3	2.0/3.3/2.3	1.3/1.0/2.3	3.0/3.0/4.0	5.0/5.0/5.0
RMSTE	3.7/3.0/1.0	2.7/3.3/3.7	1.0/1.3/2.0	2.0/2.0/4.3	5.0/5.0/4.0
CASE 3	0000 GMT November 24, 1976				
S1	4.0/3.7/ -	2.0/1.3/ -	1.7/3.3/ -	2.0/3.3/ -	5.0/5.0/ -
RMSVE	2.0/2.3/ -	3.3/1.7/ -	1.0/1.0/ -	3.0/3.7/ -	5.0/5.0/ -
RMSTE	2.7/3.3/ -	2.3/3.3/ -	1.3/1.0/ -	2.3/2.3/ -	5.0/5.0/ -
CASE 4	1200 GMT December 8, 1976.				
S1	2.0/4.0/ -	2.0/2.0/ -	1.3/1.0/ -	2.3/3.0/ -	5.0/5.0/ -
RMSVE	1.7/2.7/ -	3.3/3.0/ -	1.0/1.3/ -	3.7/2.7/ -	5.0/5.0/ -
RMSTE	2.0/2.3/ -	3.3/3.7/ -	1.7/1.7/ -	1.3/1.7/ -	5.0/5.0/ -
CASE 5	0000 GMT January 9, 1977				
S1	3.0/3.0/3.0	2.0/1.3/2.0	1.0/1.7/1.0	-	-
RMSVE	2.3/1.0/2.7	2.7/2.7/2.3	1.0/1.3/1.0	-	-
RMSTE	2.7/2.0/2.0	2.0/1.7/2.7	1.0/1.7/1.3	-	-
CASE 6	1200 GMT February 21, 1977				
S1	3.0/3.3/2.0	1.3/2.0/1.7	1.7/1.3/2.0	4.0/2.7/ -	5.0/5.0/4.0
RMSVE	2.0/2.3/1.7	2.0/2.3/3.0	1.3/1.0/1.3	3.7/3.7/ -	5.0/5.0/5.0
RMSTE	3.7/2.3/2.3	2.3/4.0/2.3	1.7/1.3/1.0	1.7/1.7/ -	5.0/5.0/4.0

Table II-5

CASE-BY-CASE
RANKING OF OBJECTIVE VERIFICATION STATISTICS FOR FOUR HEMISPHERIC MODELS AND PERSISTENCE

NORTHERN HEMISPHERE 100 MB 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
100 MB					
CASE 1					
S1	3/2/2/ -	2/4/1/ -	5/5/5/ -	4/3/3/ -	1/1/4/ -
RMSVE	1/1/1/ -	2/2/3/ -	4/5/3/ -	3/3/5/ -	4/4/2/ -
RMSTE	1/1/1/ -	2/2/2/ -	5/5/5/ -	4/4/4/ -	3/3/3/ -
CASE 2					
S1	5/3/4/ -	1/1/1/ -	3/4/2/ -	4/5/3/ -	2/2/5/ -
RMSVE	1/1/1/ -	3/5/5/ -	3/2/2/ -	5/2/4/ -	4/4/3/ -
RMSTE	2/1/1/ -	3/3/3/ -	5/5/5/ -	4/4/4/ -	1/2/2/ -
CASE 3					
S1	5/2/-/ -	1/1/-/ -	4/3/-/ -	2/4/-/ -	3/5/-/ -
RMSVE	2/1/-/ -	1/2/-/ -	4/3/-/ -	5/4/-/ -	3/5/-/ -
RMSTE	2/2/-/ -	3/3/-/ -	5/5/-/ -	4/4/-/ -	1/1/-/ -
CASE 4					
S1	1/5/-/ -	1/1/-/ -	4/3/-/ -	5/4/-/ -	3/2/-/ -
RMSVE	1/1/-/ -	3/1/-/ -	4/3/-/ -	5/4/-/ -	2/4/-/ -
RMSTE	1/1/-/ -	2/2/-/ -	5/5/-/ -	4/4/-/ -	3/3/-/ -
CASE 5					
S1	2/3/2/ -	3/2/1/ -	1/1/2/ -	-	-
RMSVE	1/1/1/ -	2/2/2/ -	3/3/2/ -	-	-
RMSTE	2/1/1/ -	1/2/2/ -	3/3/3/ -	-	-
CASE 6					
S1	1/1/2/ -	2/1/1/ -	5/3/3/ -	2/5/-/ -	5/5/4/ -
RMSVE	1/1/1/ -	2/2/2/ -	4/3/2/ -	5/4/-/ -	3/5/5/ -
RMSTE	2/2/3/ -	1/3/2/ -	5/5/4/ -	4/4/-/ -	2/1/1/ -

Table II-6

AVERAGE RANKS OF OBJECTIVE VERIFICATION STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTHERN HEMISPHERE 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
850	3.17/3.67/3.00	1.83/2.50/2.25	1.50/1.00/1.00	2.60/2.60/3.50	5.00/5.00/4.67
500	3.00/3.50/2.25	2.17/1.83/2.00	1.17/1.50/1.50	3.20/3.00/4.50	5.00/5.00/4.33
300	3.33/3.33/3.25	1.83/1.67/1.75	1.50/1.33/1.25	2.40/3.00/3.50	5.00/5.00/4.67
100	2.83/2.67/2.50	1.67/1.67/1.00	3.50/3.17/3.00	3.40/4.20/3.00	2.80/3.00/4.33
RMSVE					
850	2.67/2.00/2.00	3.00/3.00/3.25	1.17/1.00/1.00	2.60/3.00/4.00	5.00/5.00/4.67
500	2.17/2.50/2.00	2.50/2.67/2.00	1.17/1.50/1.50	3.60/2.40/4.00	5.00/5.00/4.67
300	2.00/2.33/1.75	2.83/1.67/2.50	1.00/2.00/1.75	3.40/3.60/4.00	5.00/5.00/4.67
100	1.17/1.00/1.00	2.17/2.33/3.00	3.67/2.83/2.25	4.60/3.40/4.50	3.20/4.40/3.33
RMSTE					
850	3.67/2.67/2.50	2.00/3.17/3.00	1.50/1.50/1.25	1.60/1.80/3.00	5.00/5.00/4.00
500	3.00/3.00/2.50	2.17/2.83/3.00	1.17/1.00/1.25	2.60/2.20/3.00	5.00/5.00/4.67
300	2.50/2.33/2.00	3.67/3.33/2.75	1.17/1.50/1.75	2.00/1.60/3.00	5.00/5.00/4.33
100	1.67/1.33/2.00	2.00/2.50/2.25	4.67/4.67/4.25	4.00/4.00/4.00	2.00/2.00/2.00

Table II-7

AVERAGE RANKS OF OBJECTIVE VERIFICATION STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTH AMERICA 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
1000	3.67/3.67/3.00	2.17/1.83/1.50	1.33/1.67/1.75	2.50/2.50/3.50	4.83/4.83/4.67
850	3.50/3.33/2.25	1.83/1.50/2.00	1.67/2.17/1.75	2.60/2.80/4.00	5.00/5.00/4.67
500	2.83/3.33/2.25	1.67/1.33/1.75	1.50/1.83/2.00	3.80/3.20/4.00	5.00/5.00/4.67
300	2.83/2.67/2.25	2.17/2.00/1.75	1.67/1.83/2.25	3.40/3.20/3.50	5.00/5.00/4.67
100	2.00/2.00/1.75	2.17/1.00/2.00	2.83/3.00/2.25	3.20/3.80/4.00	4.40/4.40/4.67
RMSVE					
850	2.83/2.67/1.50	1.83/2.00/2.25	1.17/1.50/1.75	3.40/3.20/4.00	5.00/5.00/4.67
500	2.50/2.83/2.00	1.83/1.50/2.25	1.00/1.67/1.75	3.80/3.40/4.00	5.00/5.00/4.67
300	2.17/2.17/2.00	2.50/2.00/2.50	1.83/1.83/1.75	3.80/3.40/3.50	4.20/4.80/4.33
100	1.33/1.17/1.25	3.00/2.67/3.75	3.33/3.33/1.75	3.40/3.40/3.50	3.40/4.20/4.00
RMSTE					
850	3.50/2.83/2.75	2.00/3.67/2.75	1.83/1.00/1.25	2.20/2.20/2.50	4.80/4.60/4.33
500	2.33/2.33/2.00	2.33/2.67/3.25	1.50/1.67/1.25	3.20/3.00/3.00	5.00/5.00/4.67
300	3.17/2.17/2.50	2.50/3.33/2.50	1.50/2.00/2.25	2.20/1.60/1.50	5.00/4.80/4.67
100	1.83/1.17/1.50	2.00/2.83/2.00	4.17/4.33/4.25	3.40/3.40/3.00	2.80/2.60/2.67
PRECIP	2.20/2.70/2.20	3.30/2.90/2.40	2.70/2.00/2.30	1.80/2.50/2.80	

Table II-8

OBJECTIVE EVALUATION

AVERAGE OF RANKS OF STATISTICS, ALL CASES, ALL TIMES

1000 - 300 mb

Statistic	9L HEM	NGM	HFM	6L PE	PER
S1	3.05	1.87	1.58	3.20	4.86
RMSVE	2.23	2.34	1.46	3.51	4.82
RMSTE	2.65	2.83	1.46	2.34	4.77
PRECIP (Ts$_p$ and Bias Combined)	2.37	2.91	2.35	2.24	

In all this concentration on relative rankings one would not want to loose sight of the actual error scores - Tables II-9 and II-10 are presented to avoid that pitfall. They contain the values of the error quantities averaged over the cases as functions of pressure and forecast time for the Northern Hemisphere and North American networks, respectively. As an aid in the interpretation of tables II-9 and II-10 companion tables II-11 and II-12 show the ranking of the average scores, as before, on a one to five basis. The conclusions to be drawn from these rankings of the average scores are the same as those drawn from the average ranks of the individual scores: HFM best in the troposphere, NGM second best or perhaps tied for first over the North American area while, stratospherically, the HFM makes a poor showing and the 9LH and NGM vie for top honors. The most remarkable thing about Tables II-9 and II-10 is how close together the scores are (except for tropospheric persistence) - the various forecasts were not wildly divergent and the concentration on ranking served to emphasize the small differences. But what is equally remarkable, as has been emphasized, is the consistency of the small differences - there is not question that the HFM is best, albeit by a small margin.

The precipitation scores on Table II-10 are not the average of the individual case scores but (properly) the scores for the ensemble of cases. Again, as with the ranked precipitation scores, there is no clear winner in this category.

2. Subjective Evaluation

Here the rather considerable amount of discussion, inter-map comparisons, time spent in seminar presentation by individual evaluators, and tabulations of ranks by categories in the subjective evaluation forms (Appendix I) have been reduced to a single table, II-13. In this we have, in the top half, the rankings averaged over all categories of the subjective evaluation form, over all cases, and all times by individual forecaster. Because of the difficulties with precipitation forecasting the rankings related only to precipitation categories were averaged to see if the subjective evaluations were any more definitive than the objective scores. Obviously, they were not.

The unanimity among jurors of the all category rankings and their agreement (as to relative merit of the models) with the average of the objective rankings (the jurors' attention was centered on the North American area where the NGM finer mesh insert was located) is both heartening and a firm confirmation of the objective score results.

Table II-9

AVERAGE VALUES OF OBJECTIVE VERIFICATION STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTHERN HEMISPHERE 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
850	53.2/59.9/71.9	49.7/55.5/66.9	49.4/52.7/62.0	51.9/55.7/79.6	69.2/79.5/84.7
500	45.5/49.4/64.1	44.5/46.0/62.0	43.2/44.3/57.7	47.3/47.8/75.4	60.5/71.7/78.6
300	44.6/49.9/63.3	43.9/48.0/59.4	42.1/46.9/57.2	45.9/51.7/71.2	60.1/71.7/79.7
100	66.1/65.3/73.7	66.1/63.2/68.6	67.1/66.1/70.9	68.5/68.9/80.2	66.5/66.8/79.1
RMSVE (m/s)					
850	6.7/ 7.9/ 9.0	7.1/ 8.2/ 9.7	6.3/ 7.7/ 8.5	6.8/ 8.0/ 9.2	9.4/11.5/11.7
500	8.4/11.2/13.7	8.5/10.5/13.9	7.9/10.1/13.2	8.9/11.1/14.6	14.5/19.4/19.5
300	11.9/15.4/18.4	12.3/14.9/18.9	11.2/14.5/18.2	12.9/15.8/21.0	21.2/27.6/27.4
100	7.5/ 8.7/ 9.8	8.4/10.0/12.9	9.8/10.8/12.3	10.2/11.7/13.2	8.6/12.6/13.1
RMSTE (°C)					
850	3.3/ 4.0/ 5.4	2.9/ 4.3/ 5.5	2.9/ 3.7/ 4.5	2.9/ 3.5/ 5.1	4.5/ 6.0/ 6.9
500	2.2/ 3.1/ 4.2	2.2/ 3.0/ 4.3	2.0/ 2.7/ 3.5	2.2/ 2.9/ 3.9	4.2/ 5.7/ 5.9
300	2.6/ 3.3/ 3.7	2.8/ 3.5/ 3.9	2.3/ 3.1/ 3.7	2.4/ 3.1/ 3.9	3.4/ 4.4/ 4.6
100	2.9/ 3.3/ 4.1	3.2/ 4.1/ 4.9	5.2/ 6.7/ 8.7	4.8/ 5.9/ 5.9	3.0/ 3.8/ 4.2

Table II-10

AVERAGE VALUES OF OBJECTIVE VERIFICATION STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTH AMERICA 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
1000	53.7/66.5/66.3	46.9/56.3/59.0	46.2/54.7/57.1	49.8/59.3/80.7	78.9/96.8/84.3
850	47.9/57.6/67.6	43.4/47.5/57.9	42.8/49.8/57.5	44.8/49.9/72.0	67.6/85.3/90.9
500	33.4/42.1/49.7	31.9/35.4/45.7	31.5/36.5/46.6	35.8/41.1/57.0	55.7/67.7/74.8
300	32.5/38.0/47.8	30.9/34.6/44.6	30.6/34.4/46.5	34.4/39.5/54.9	55.8/67.3/74.6
100	47.1/47.3/58.5	46.5/45.8/55.9	49.1/49.5/56.2	51.5/53.5/67.6	52.6/55.8/70.9
RMSVE (m/s)					
850	7.3/ 9.5/ 8.9	6.7/ 9.2/ 9.6	6.5/ 8.9/ 8.6	6.9/ 9.2/ 9.7	11.5/14.9/13.1
500	8.9/12.5/13.8	8.7/10.5/13.3	8.1/11.1/13.7	9.3/12.0/13.4	18.1/23.9/20.1
300	12.5/15.8/19.3	12.4/15.1/18.5	12.0/15.2/19.9	13.5/16.6/21.8	26.4/33.2/31.8
100	7.0/ 8.4/10.5	9.2/11.6/15.6	9.5/10.7/13.6	9.6/10.8/13.1	9.9/12.8/16.0
RMSTE (°C)					
850	4.1/ 4.9/ 6.4	3.4/ 5.2/ 5.7	3.3/ 4.0/ 4.5	3.2/ 4.4/ 5.6	6.4/ 8.5/ 8.4
500	2.2/ 3.3/ 4.5	2.2/ 3.3/ 5.2	2.0/ 3.0/ 4.2	2.3/ 3.2/ 4.6	4.9/ 6.7/ 7.7
300	2.9/ 3.0/ 3.8	2.8/ 3.2/ 3.9	2.5/ 2.9/ 3.7	2.6/ 2.9/ 3.3	3.7/ 4.5/ 6.5
100	2.8/ 3.4/ 4.5	3.6/ 4.7/ 4.8	5.1/ 6.3/ 8.3	4.6/ 5.2/ 6.7	3.5/ 4.4/ 5.7
Ts_p	40.4/23.7/17.2	36.2/25.3/25.9	41.1/30.9/25.0	46.7/25.1/12.0	
BIAS	80.0/109 /76.0	170 /227 /224	156 /201 /127	128 /149 /81.0	

Table II-11

RANKING OF AVERAGE VALUES OF OBJECTIVE STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTHERN HEMISPHERE 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
850	4/4/3	2/2/2	1/1/1	3/3/4	5/5/5
500	3/4/3	2/2/2	1/1/1	4/3/4	5/5/5
300	3/3/3	2/2/2	1/1/1	4/4/4	5/5/5
100	1/2/3	1/1/1	4/3/2	5/5/5	3/4/4
RMSVE					
850	2/2/2	4/4/4	1/1/1	3/3/3	5/5/5
500	2/4/2	3/2/3	1/1/1	4/3/4	5/5/5
300	2/3/2	3/2/3	1/1/1	4/4/4	5/5/5
100	1/1/1	2/2/3	4/3/2	5/4/5	3/5/4
RMSTE					
850	4/3/3	1/4/4	1/2/1	1/1/2	5/5/5
500	2/4/3	2/3/4	1/1/1	2/2/2	5/5/5
300	3/3/1	4/4/3	1/1/1	2/1/3	5/5/5
100	1/1/1	3/3/3	5/5/5	4/4/4	2/2/2

Table II-12

RANKING OF AVERAGE VALUES OF OBJECTIVE STATISTICS FOR ALL CASES FOR FOUR HEMISPHERIC FORECAST MODELS AND PERSISTENCE

NORTH AMERICA 24/48/84 HRS

	9L HEM	NGM	HFM	6L PE	PER
S1					
1000	4/4/3	2/2/2	1/1/1	3/3/4	5/5/5
850	4/4/3	2/1/2	1/2/1	3/3/4	5/5/5
500	3/4/3	2/1/1	1/2/2	4/3/4	5/5/5
300	3/3/3	2/2/1	1/1/2	4/4/4	5/5/5
100	2/2/3	1/1/1	3/3/2	4/4/4	5/5/5
RMSVE					
850	4/4/2	2/2/3	1/1/1	3/2/4	5/5/5
500	3/4/4	2/1/1	1/2/3	4/3/2	5/5/5
300	3/3/2	2/1/1	1/2/3	4/4/4	5/5/5
100	1/1/1	2/4/4	3/2/3	4/3/2	5/5/5
RMSTE					
850	4/3/4	3/4/3	2/1/1	1/2/2	5/5/5
500	2/3/2	2/3/4	1/1/1	4/2/3	5/5/5
300	4/3/3	3/4/4	1/1/2	2/1/1	5/5/5
100	1/1/1	3/3/2	5/5/5	4/4/4	2/2/3

Table II-13

SUBJECTIVE EVALUATION

AVERAGE OF RELATIVE RANKING OF FOUR
NMC HEMISPHERIC MODELS FOR ALL CASES, ALL TIMES
ALL CATEGORIES

JUROR	9L HEM	NGM	HFM	6L PE
HKS	3.8	2.0	1.1	3.7
JOC	4.0	2.0	1.3	3.2
HEB	3.5	2.5	1.1	3.5
RMM	3.9	2.6	1.0	3.0
DAO	3.8	2.4	1.0	3.4
Mean	3.8	2.3	1.1	3.1
AVERAGE OF RELATIVE RANKING PRECIPITATION ONLY				
HKS	2.8	3.2	2.7	2.7
JOC	2.3	2.9	2.2	1.8
HEB	2.8	2.4	2.3	2.2
RMM	2.3	2.3	2.8	2.7
DAO	1.7	2.8	2.8	2.8
Mean	2.4	2.7	2.6	2.4

A few remarks are appropriate as a summary of the sence of the subjective evaluations of the forecasts - the "map discussions" rather than the subjective rankings of the categories. First as to the precipitation forecasts. They were, in general, rather a disappointment from all the models. There was no clear cut improvement over the 6LPE from any of the finer mesh models (this is obviously reflected in the inconclusive ranking scores) and it was not a case of good forecasts not being improved but rather a case of not so good forecasts staying that way. Some of the objective precipitation scores in Table II-10 are, in absolute terms, rather good. This does not agree with the forecaster's evaluations - the utility of the computer forecasts as guidance to them is their basic measure of worth - and we can only conclude that either the results of Table II-10 are fortuitious or, more likely, the particular objective scoring method is not a good measure of precipitation forecast utility.

A happier story can be told relating to some of the synoptic features that have long troubled the 6LPE forecasts. The first of these is the "cross contour flow" pattern - all of the models cured the problem all but completely; some of the forecasts show a little residual problem but the improvement is so substantial that one hardly notices that it is not complete. With a slightly lesser degree of enthusiasm the same may be said of the inconsistencies between height and vorticity patterns: They have been substantially reduced in severity and frequency, although from time to time the problem does recur but with less severity. The improvement in the "locked-in-error" appears to be substantial but not total; the impression is that about 50% of the displacement error has been taken care of. These improvements were not the same in all the models; the variations from model to model in these and other respects (placement and depth of surface lows, translation of troughs, strength of jets, etc., etc.,) are the basis for the subjective rankings which brought the HFM to the top. Specific examples of these synoptic feature improvements are shown below in connection with the further testing of the best-of-the-three winner vs the 6LPE.

3. Special Aspects: High Altitude Error Structure, Energetics
(Summary Report Prepared by Upper Air Branch, DD, NMC)

High altitude error structure. Model performance in the stratosphere is of interest for various reasons, including the ability to forecast flight conditions for supersonic aircraft. Graphs of all case average forecast error (here verified against analyses) versus altitude show a reversal in the ranking of the models above 300 mb as do the statistics of Tables II-9 and II-10. Referring to Fig. II-2, HFM has generally smallest rms height error at 1000-300 mb, except over North America, where the NGM has least error at 500-300 mb.

In striking contrast, HFM is "worst" (greatest error) at 24 and 48 hours at 100 mb. The 9LH improves dramatically above 300 mb, having least error at 100 mb. Under certain conditions, however, 9LH was observed to have large errors near the pole at 100 mb.

In terms of rms temperature error, Fig. II-3, it is noteworthy that (1) 9LH has least error both at 100 mb and near the ground; and that (2) the models are much better than persistence only at 500-300 mb. Over North America , 9LH was clearly better than persistence at 100 mb (not shown).

The statistical results in Figures II-2 and II-3 are based on comparisons of forecast against analyses. Additional comparisons with radiation-corrected station data over North America and over a Transatlantic SST route showed the same ranking at 100 mb as above.

Energetics. As an additional diagnostic aid, the forecasts for the 9LH, NGM, and HFM models were analyzed by means of the NMC atmospheric energy program (Miller, et. al., 1975) and compared to previous generalized results for the 6-layer model. The 6LPE energetics were obtained from Hauser and Miller, 1978, and represent typical 6LPE energy behavior, not the actual energetics of the model for the test cases. All computations were done over the domain 850-200 mb and 20N-pole for the forecast hours 12, 24, 48, and 84 when available, for the four days 8/23/75, 11/24/76, 1/9/77, and 2/21/77.

The various results are sketched in Fig. II-4. They show the energy verification ratios, defined as (forecast-verification)/verification, as a function of forecast hour, for the zonal available potential energy (AZ, top left), zonal kinetic energy (KZ, bottom left), eddy available potential energy (AE, top right) and, eddy kinetic energy (KE, bottom right).

The points marked 6LPE are the averages of the four monthly mean values given by Hauser and Miller. The other points for the other models are the averages for the 4 days cited above. In summary, it appears that both the HFM and NGM models behave energetically in an overall similar fashion to the 6LPE with the possible exception of the KE term. The 9LH, on the other hand, shows some interesting dispersions from the other models. The AZ term holds at about the same level throughout the forecast period, suggesting that the radiation physics may be better for this model than the others, but the eddy terms, AE and KE, lose energy at a much greater rate.

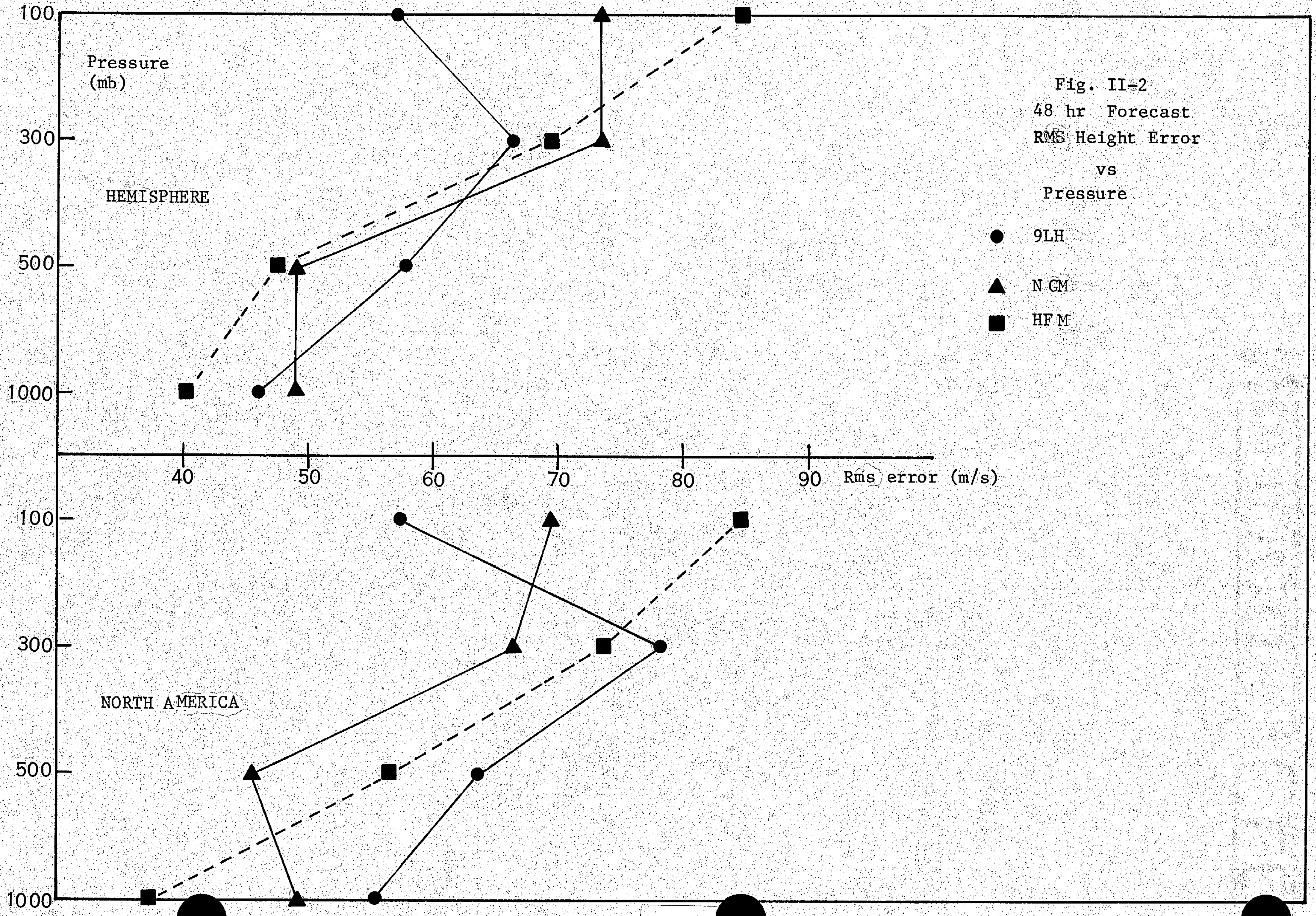


Fig. II-2
48 hr Forecast
RMS Height Error
vs
Pressure

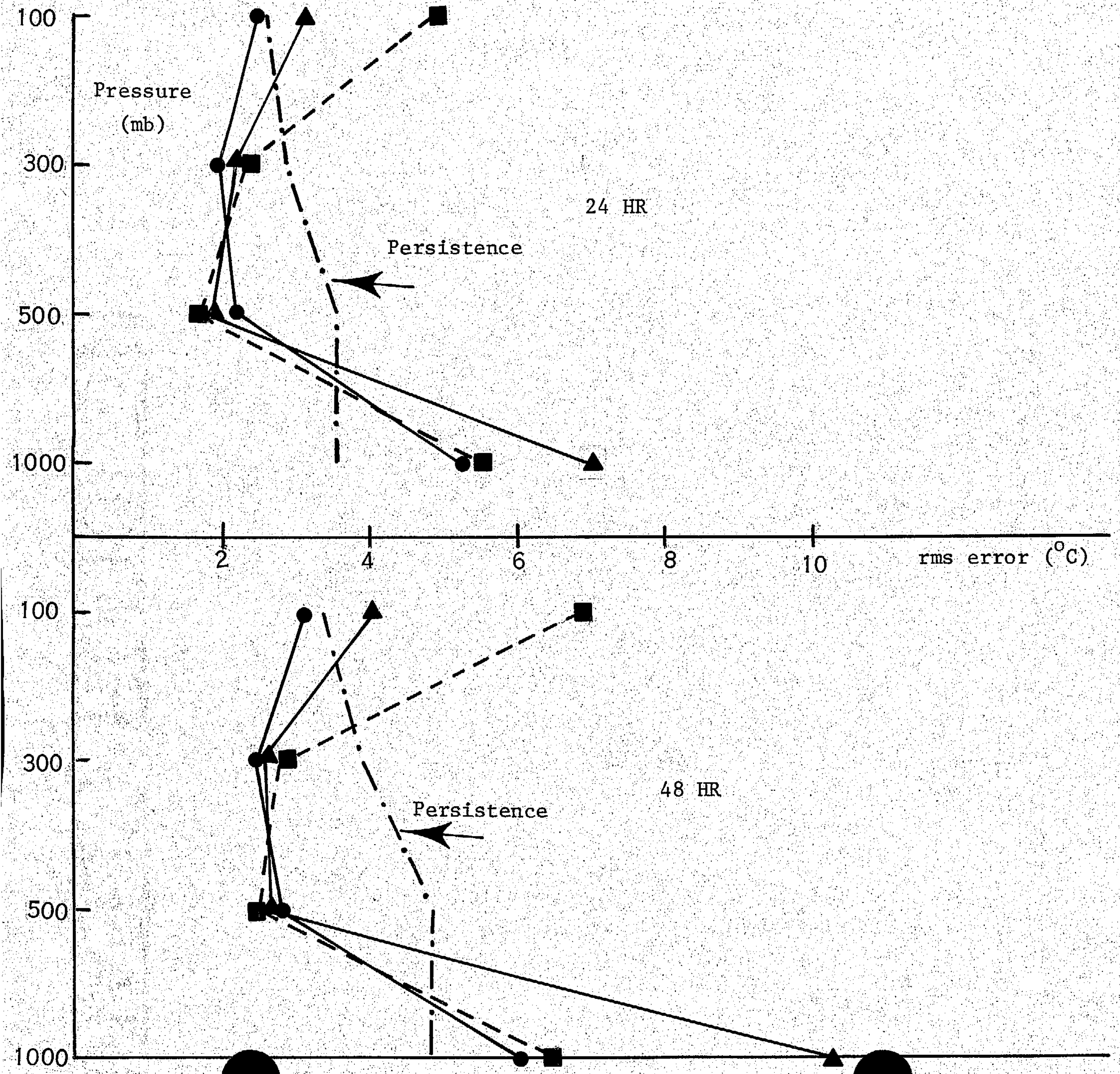


Fig. II-3

24 hr and 48 hr forecast

RMS Temperature
Error
vs
Pressure
(NORTHERN HEMISPHERE)

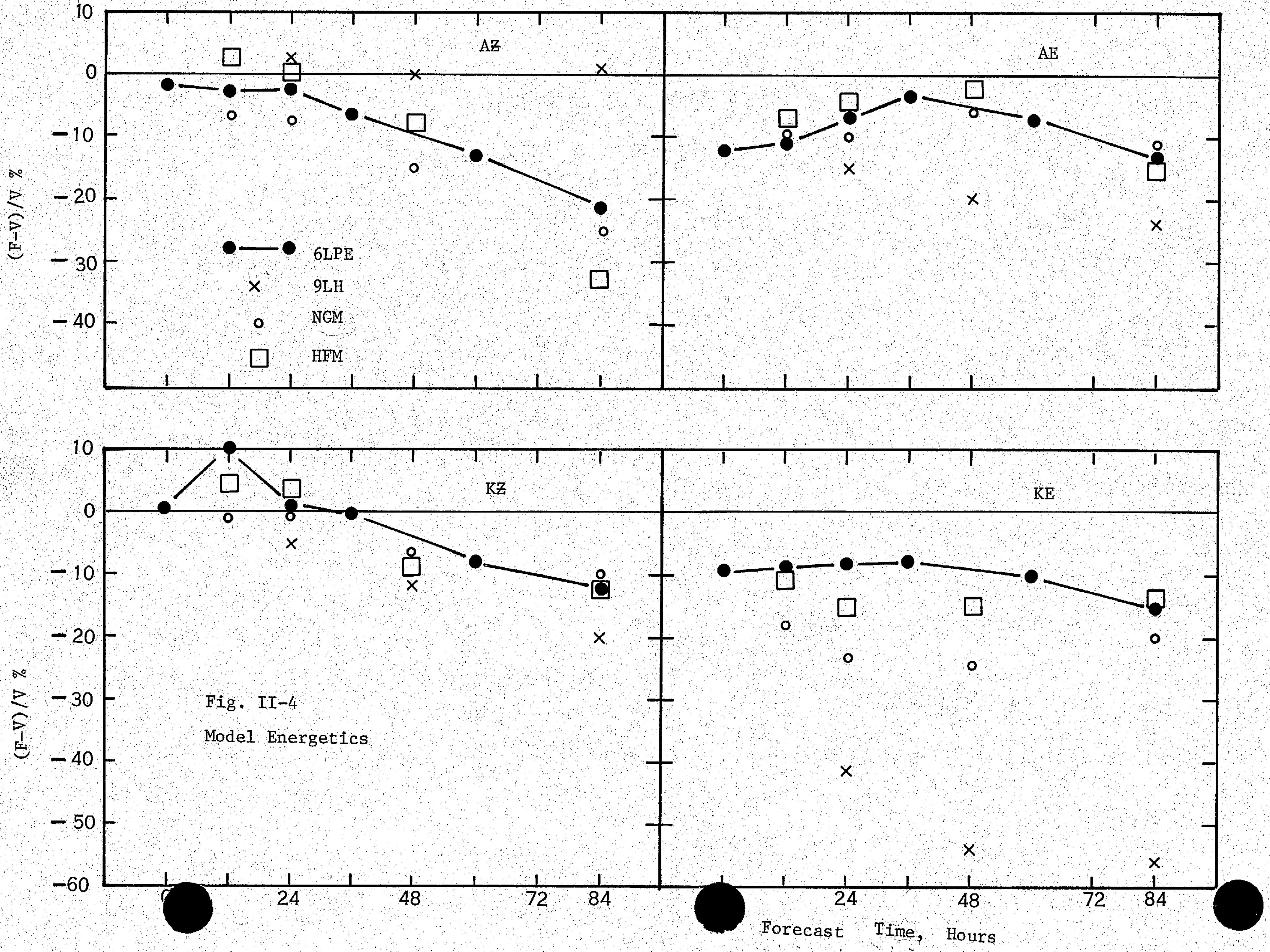


Fig. II-4

Model Energetics

Although the original plan had called for some 24 additional pairs of forecasts (beyond the 6 of the four model run-off tests) we felt that since the 7LPE differed from the 6LPE only in ways that should, _à priori_, lead to improvements (mainly increased resolution), such extensive testing would not be necessary. Ten cases were deemed adequate. Table III-1 lists the initial times and some of the salient features of the selected cases.

Table III-1 Ten Case Test Selections

12Z 09 Jan 75	Major Winter Storm, locked-in error.
00Z 11 Jan 75	Ditto
00Z 21 Jan 75	Midwest Storm, locked-in error.
00Z 05 Feb 76	All from the Data Systems Test (DST) for which extra data were available
00Z 09 Feb 76	
00Z 17 Feb 76	
12Z 05 Dec 76	Good 6LPE forecast in Pacific. Extensive precipitation areas.
12Z 05 Mar 77	Locked-in error
00Z 01 Jul 77	Summertime Convection
00Z 08 Jul 77	Ditto

As with the four model run-off, the cases were selected partially on the basis of availability with the implicit assumption (or hope) that the operational uses for which the model is put would be exemplified in these cases.

The subjective and objective evaluations of the forecasts proceeded essentially as with the four model tests, with some differences. The same sets of maps were prepared, the same evaluation questions asked of the forecasters, and the same objective statistics computed. There was no effort to disguise the models (it would have been futile) and rather than give all 10 forecasts sets to each of the 5 jurors (rather an overwhelming prospect) each juror received two different cases and was called upon to discuss his evaluations in a seminar setting.

As ever, not all the cases could be completed to 84 hours: None of the 12Z runs went past 48 hours operationally; the tape for 00Z 11 Jan 75 was destroyed before the 84 hour objective statistics could be computed; the same was true for the 24 hr forecast from 00Z 9 Feb 76; and finally 00Z 1 July77 received no subjective evaluation at all. (The evaluator announced in his seminar that 1 July was a "very uninteresting case" and turned his full attention to the other case he was responsible for. The evaluator did suggest that both forecasts, 7LPE _vs_ 6LPE from 1 July, were very similar and had he looked closely they probably would have been tied in the various categories).

B. Results

1. Objective Statistics

With only two models to compare we can do away with extensive considerations of relative rankings and turn directly to the summary scores. Table III-2 shows the average value of the various statistics for all cases for the North American station observation set, broken out by pressure. A quick perusal shows that for every statistic, level and time (save only 48 and 84 hr 100 mb S1, and 84 hr 500 mb temperature) the 7LPE bests or equals the 6LPE, frequently by appreciable amounts. The precipitation scores, however, do not share in this tendency - as with the four model test precipitation forecasting is not helped much by the 7LPE at least as far as these objective scores are concerned.

It is instructive to compare Table III-2 and Table II-10 with particular reference to the 6LPE and the 7LPE/HEM scores in the respective tables. The two sets of 6LPE scores look quite similar in character which suggests, at any rate, that the 10 cases of Table III-2 and the 6 of Table II-10 were not unique in any way - they were both representative samples. The same may be said of the 7LPE/HEM statistics except at 100 mb. At that level the anticipated improvement from the seventh level seems to be realized, in the winds and temperatures at any rate. The S1 scores (and the rms height errors, not shown here) do not show any improvement at 100 mb with the seventh level. This remains as something of a puzzle.

Another comparison of interest is the 7LPE, at 100 mb, relative to persistence. Persistence "forecasts" were not verified for the 10 case, two model tests but, to the extent that the 6 and 10 cases are statistically similar, it would seem that the 7LPE at 100 mb is capable of beating persistence where the 6LPE (still) is not. This is a somewhat tentative conclusion, however, and it does not apply to the S1 scores with any assurance. Only time and further routine verifications will tell.

Other statistics, verifications with respect to other sets of observations, and with respect to analyses all agree in showing the 7LPE model to be clearly superior to the 6LPE.

2. Subjective Evaluations

With only two models to select from, the problem of relative ranking reduces to simply voting for either model as better (or stating that they are tied) and the summary table, Table III-3, is just a tally of those votes. The Table presents a) the total votes for each model (and for ties) for all categories of judgment (Appendix I), all cases, and all times, and b) the votes for the precipitation related categories only.

Table III-2

Average Value of Objective Verification Statistics for All Cases

NORTH AMERICA 24/48/84 HRS

	7LPE	6LPE
S1		
850	48.2/47.4/61.4	52.1/56.0/70.2
500	34.7/36.9/46.2	34.7/41.1/52.1
300	30.4/35.5/49.6	31.0/39.8/51.7
100	57.8/61.1/59.9	59.6/59.8/59.4
RMSVE(m/s)		
850	6.6/6.1/10.6	7.4/8.7/11.6
500	8.6/10.0/16.2	9.3/12.5/17.4
300	11.0/15.1/22.0	12.6/17.9/24.6
100	7.7/9.5/11.5	10.1/12.8/16.8
RMSTE(°c)		
850	2.8/3.7/5.1	3.1/4.4/5.1
500	2.1/2.7/4.2	2.3/3.2/3.8
300	2.1/2.4/3.5	2.3/2.8/4.2
100	3.7/3.7/4.9	4.1/5.8/6.9
T_{sp}	41/38/22	42/32/22
BIAS	147/150/185	147/159/222

Table III-3

Subjective Choice (Total Votes) all categories,
all cases, all times

	7LPE	TIE	6LPE
VOTES	135	52	46

Precip Categories Only

VOTES	37	30	27

The 7LPE is obviously a clear winner by almost 3 to 1 (5 to 1 if you subtract the precipitation categories votes from the totals for all categories). This confirms the judgments based upon the objective statistics and agrees with the 6LPE vs HFM results of the four model run-off tests. The general lack of improvement in precipitation forecast is also reflected in Table III-3 - as before in the map discussions of the forecasts disappointment was felt that nothing much had happened to better the precipitation forecasts.

3. Stratospheric Evaluations
(Summary Report by Upper Air Branch, DD, NMC)

A. Objective Statistics

Root mean square (rms) differences between forecast fields and verifying analyses (at gridpoints) served as the basic method for comparing the models. Given an rms fit for each case (e.g., x_i) overall rms errors were calculated using the formula $\frac{\Sigma (x_i)^2}{n}$, where n is the number cases.

As expected, rms errors were different in magnitude over various areas of the Northern Hemisphere and for various forecast times (24 and 48 hours). However, error profiles and thus <u>relative</u> forecast quality remained consistent over all areas and verification times. Studies were thus limited to area 1 (195 gridpoints over North America) and area 6 (1977 gridpoints over Northern Hemisphere octagon).

Note that in Figure III-1 (area 1, 24 hours) the 7LPE fits the verification better than the 6LPE for all levels and parameters (9-case average). The 7LPE does better than persistence in all categories except 100 mb temperatures. It is important to note that previous studies have shown that the 6LPE does not beat persistence in any parameter at 100 mb. Figure III-1 is consistent with those results and implies that the 7LPE wind forecasts <u>are</u> superior to persistence. The 7LPE would thus provide increased skill for such applications as supersonic transport forecasts. Somewhere above 200 mb persistence overcomes both forecast models in predicting temperature.

B. Subjective Studies

Subjective evaluation results from the 10-case comparison between the 7LPE and the 6LPE (course mesh) at 100 mb are shown in Table III-4. Each case was studied independently by two evaluators. The two models were rated by parameter (height, temperature, and wind fields) using combined information from the 24-hour and 48-hour forecasts and verifications. Each evaluator gave one one vote to the model of his choice with the option of allowing ½ vote for each model, if tied.

Table III-4 shows that over the 10-case summary the 7LPE was judged to be better than the 6LPE, although for individual cases the 6-layer was found to be superior in certain parameters. In all cases 7LPE temperature forecasts were rated superior, primarily because of a persistent 6LPE cold bias at low- and mid-latitudes. In six of the ten cases the height and wind fields of the 7-layer were rated superior.

4. ATA Winds

One of the principal uses of the 6LPE is the preparation of high level wind forecasts for airline flight planning purposes. These are known as the ATA (for Air Transport Association) winds. In addition, amendments to the machine forecasts are readied manually from time to time when appropriate. It is the purpose of this section to compare the short range, mid-level wind forecast from the 6LPE and the 7LPE and to suggest revisions in the NMC ATA wind amendment techniques for the 7LPE. Data used in this study consist of a difference chart created by subtracting the 6LPE and 7LPE 24 hour 300 mb wind speed forecasts from the NMC final analyses of the 300 mb wind speeds at each grid point on the NMC 65x65 grid. Wind speed forecast errors of at least 20 knots are considered significant errors. Comparison points are the gridpoints at the NMC 65x65 grid at all latitudes between the Asian Pacific Coast eastward to 40 degrees East longitude. The comparison area roughly represents the region over which ATA wind amendments are issued.

Table III-5, which includes an extra forecast for 9 Sept 77 made for this comparison, shows the number of points where the 7LPE and the 6LPE overforecast and underforecast the wind speed by at least 20 knots. The Table shows that 1) the 7LPE underforecast the wind speed at about the same number of points as did the 6LPE, 2) the 7LPE overforecast the wind speed at about one half the number of points as did the 6LPE, and 3) the 7LPE wind speed forecasts have a significant error at approximately 80% of the number of 6LPE error points. The ratio of points where the wind speeds were significantly underforecast to points where the wind speeds were significantly over forecast is 3:1 for the 7LPE and 3:2 for the 6LPE. It is concluded from the table that revisions of the ATA wind forecast amendment procedures for the 7LPE should include the reduction of areas for wind speed overforecasting corrections while leaving the wind speed underforecasting corrections approximately the same size.

Fig. III-1

24 HR Forecast Verifications

vs. Analysis for North

American Area

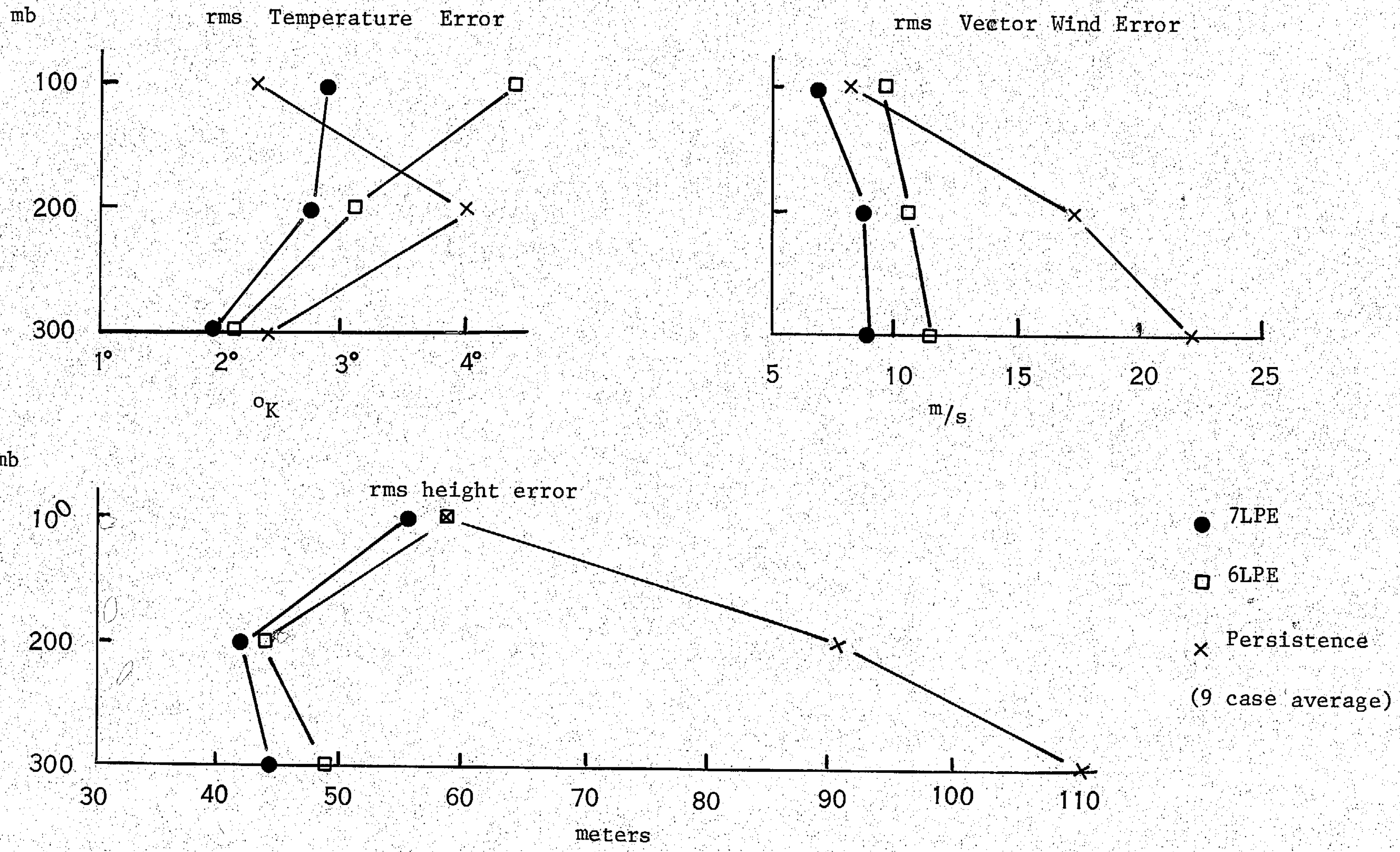


Table III-4

Subjective Rating of 24 and 48 HR <u>100 mb</u>

MODEL	FORECAST FIELDS					
	6LPE			7LPE		
CASE	H	T	W	H	T	W
9 Jan 75	2	0	2	0	2	0
11 Jan 75	½	½	1	1½	1½	1
21 Feb 75	½	½	½	1½	1½	1½
5 Feb 76	0	0	0	2	2	2
9 Feb 76	0	0	0	2	2	2
17 Feb 76	0	0	0	2	2	2
5 Dec 76	1	0	1	1	2	1
5 Mar 77	1½	0	0	½	2	2
1 Jul 77	1	0	1	1	2	1
8 Jul 77	½	½	½	1½	1½	1½
TOTAL	7	1½	6	13	18½	14

Table III-5

DATE:	Jan 975	Jan 1175	Feb21 75	Feb 576	Feb 976	Feb 1776	Dec 576	Mar 577	Jul 177	Jul 877	Sep 977	Totals
7LPE Underforecast Points	159	201	178	220	211	161	78	158	48	70	69	1553
6LPE Underforecast Points	199	242	163	229	176	158	64	126	41	41	74	1513
7LPE Overforecast Points	71	78	44	52	42	67	33	43	5	15	16	466
6LPE Overforecast Points	126	120	112	89	111	114	83	120	38	34	56	1003
7LPE Total Error Points	230	279	222	272	253	228	111	201	53	85	85	2019
6LPE Total Error Points	325	362	275	318	287	272	147	246	79	75	130	2516

Number of Points where the difference between the 24 hour forecast wind speed and the analyzed verification wind speed is at least 29 knots.

Study of a number of particular cases leads to these general conclusions:

The 7LPE forecast trough speeds are about the same or faster than in the 6LPE. Forecast wind speed errors, because of slow forecast trough speeds in the 6LPE, should be less for the 7LPE than for the 6LPE. The 7LPE forecast wind speeds are often less than in the 6LPE. Overforecast wind speed errors should be less for the 7LPE than for the 6LPE. The 7LPE shows no improvement over the 6LPE with respect to amount of wind speed underforecasting errors, but shows dramatic improvement with respect to wind speed overforecasting errors. A general rule in revising the ATA wind amendment techniques for the 7LPE is to reduced or eliminate the wind speed overforecasting correction and keep the wind speed underforecasting correction about the same as for the 6LPE.

IV. Sample Comparisons

It would be a shame not to include at least a small selection of examples of the benefits to be gained by using the 7LPE over the 6LPE.

The first of these is an example of "cross contour flow". Fig IV-1 shows an admittedly rather extreme example - the map is a 36-hr 300 mb height and isotach forecast from the 6LPE over the Central Pacific. Clearly there is a problem. Fig. IV-2 is the verifying analysis with a respectable looking jet blowing zonally across the Pacific.

There is, in the analysis, only the slightest suggestion of a southward dip of the isotach maxima at 155° West and no trace of the wavy pattern in the winds further west. This wavyness was another of the 6LPE's problems, usually seen as inconsistencies between the height and vorticity patterns. Fig. IV-3 is the 7LPE 36 hr forecast. Clearly, the cross contour flow problem has been cured as have most of the height/wind inconsistencies. Further examples could be adduced from the 16 other available cases which show the same degree of improvement.

The second example is a phenomenon in the forecasts that has long been, and continues to be, a major preoccupation of NMC's modelers - the "locked-in error." Figures IV-4, a, b, c, d, e, are a sequence of analysis maps at 12 hour intervals for the sea level pressure and 500-1000 mb thickness. The forecast, for which these are the verifying maps, typifies the problem. Consider first the analysis sequence. Fig. IV-4a for 12Z/9 Jan 75 shows a respectable low over the Colorado-Kansas border. Half a day later, Fig. b, the low had moved to central Texas with little change (in this machine analysis, at any rate) of central pressure. This was about as far south as the low got; 12 hours later, Fig. c, it had deepened somewhat and traveled north eastward to about Joplin, Mo. The next 12 hours of the sequence, Map d, for 00Z/11 Jan 75, shows substantial development and rapid northward motion to Prairie du Chien, Wisconsin. By now we have a major winter storm, of course, for which accurate forecasts have considerable economic value. Finally, by 12Z 11 Jan 1975, Fig. e, we have a huge mid-continent storm centered over Duluth.

For completeness, the corresponding 500 mb height and vorticity charts are presented as Fig. IV-5, a-3. They show about what one would expect, ending with an occuluded system.

The 6-layer coarse mesh forecast through 48 hours in the Fig. IV-6, a-c, (we are not bothering with the 12 hour forecast). The 24 hour forecast, Fig. a, is not too bad. Differences in detail can be seen to be sure, but the low in question is in pretty good shape. It is in the second 24 hours of the forecast that the "locking-in" takes place. At 36 hours, Fig. IV-6b, model has managed to move the low only to central Missouri and has filled it one millibar. It doesn't take much effort to see what would become of the St. Louis 36 hour wind and temperature forecast, say, if it were based entirely on the 6-layer model forecast. Finally, by 48 hours, Fig. IV-6c, things have come to a pretty pass. To be sure the model makes a valient attempt at a Lake Superior low, but also produces an utterly spurious low in Northern Louisiana. What was true for St. Louis at 36 hours goes double for Chicago at 48 hours.

The 500 mb forecast Fig. IV-7, a-c shows, in the 48 hour portion, the same lag of the model behind the atmosphere. The major trough and vorticity center are both in southern Louisiana, when they should be in Wisconsin. An analysis of the 500 mb height error would show a huge error doublet: 300 meters too low over Louisiana, 300 meters too high over N.E. Michigan. This doublet has been suggested as the pattern defining what is meant by a "locked-in-low".

We can turn to the 7LPE model forecasts with a certain sense of relief. Fig IV-8, a-c shows the 24, 36, 48 hour sea level forecast maps and it's obvious we are in much better shape. The 24 hour forecast is just about as good as you could hope for. The 36 hour is a little slow and not deep enough but its a vast improvement over the 6LPE version. Finally, at 48 hours the 7LPE triumph is nearly complete - the forecast low is over N.E. Wisconsin rather than Duluth, but there is no trace of the spurious Louisiana low. Granted there is no high over Louisiana and the gradient over the midwest is not as strong as nature but the improvement is, to say the least, striking. The Chicago forecaster would doubtless concur. Also a rather strange collection of small highs and lows, generated by the 6LPE over the West and Southwest at 48 hours, has been replaced by a considerably more realistic forecast pattern.

For 500 mb, Fig. IV-9, a-c, analogous statements could be proffered. At 48 hours, particularly, the low center and vorticity maximum are somewhat laggard - the locked in error is not cured completely - but the improvement over the 6LPE is again impressive. The 500 mb height error doublet still exists - 180 meters too low over Mississippi, 240 meters too high over N.W. Michigan - giving further indication of the continued presence of the locked in error. Reducing the truncation error helps but is, seemingly, not the whole story.

We have given an example of the 7LPE benefits for the Pacific and Central Regions (we would have shown some cross contour flow over the Rockies for the Western region except that the Pacific one was more dramatic). The third example is for the Southern and Eastern regions. Fig. IV-10, a-c are the observed maps at 24 hour intervals from 12Z/5 Mar 77. The main item of interest is the low moving at a steady pace from the western Gulf of Mexico to a position east of Cape Hatteras. Nothing terribly dramatic - rather routine looking as a matter of act. The corresponding 500 mb charts are in Fig. IV-11, a-c. The system doesn't appear to have very extensive upper air support. The 6LPE forecast series Fig. IV-12, a-b and IV-13 a-b, (the 24 and 48 hour forecasts) obviously leave something to be desired, particularly at 48 hours. What should be a Cape Hatteras low is a clump of three lows over Florida; elsewhere, there is too strong a surface trough over Colorado. The 500 mb vorticity pattern over Texas and Mexico is rather exotic too. The 7LPE forecasts, Fig. IV-14, a-b, and Fig. IV-15, a-b, are obviously much closer to the mark. The Hatteras low is in place, the Colorado Trough is greatly reduced and, at 500 mb, the height and vorticity patterns are much more pleasing to the eye, not to mention being more correct.

These various case examples and, more particularly, the overall statistical results described above are what lead to the decision to implement the 7LPE into NMC's operations. This took place on 12Z/19 Jan 78 amid much rejoicing.

Since the better part of a year has gone by, we are able to show one more piece of (post facto) evidence of the forecast improvement with the 7LPE model. Since June of 1977 NMC has been verifying the operational model forecasts against observations from a network of 102 quasi-uniformly distributed Northern Hemisphere radiosonde stations. This is in cooperation with the Navy and Air Force who are verifying their numerical forecasts against the same observation set. Fig. IV-16 shows the monthly root-mean-square wind vector errors (in meters/sec) for the 48 hour 6LPE or 7LPE forecasts at 500 mb.

The time trace of the rms error shows, obviously, a clear discontinuity in January 1978. (The two January values are the averages for days 1-19 and 20-31, the 6LPE and 7LPE portions of the month, separately.) Where the verification periods overlap, June-September, the step improvement of the winter time model change seems to be confirmed. Also plotted are the U.S. Air Force Global Weather Center forecast model verifications against the same set of observations. This serves as a sort of control as the USAF model has not undergone any major change. Most striking is the closeness of the Air Force model and the 6LPE, and the closely parallel seasonal variation of the Air Force and both 6 and 7LPE models. The consistent improvement of the 7LPE over the Air Force "control" model is, of course, evident.

Other statistical error measures, S1 scores, rms heights and temperatures, all show a similar (not in all cases as large) improvement in the historical record. With the 7LPE model, everything got better except the precipitation forecasts.

References

Brown, J. A., Jr., and K. A. Campana, 1978: An economical time-differencing system for numerical weather prediction. Mon. Wea. Rev., 106, 1125-1136.

Hauser, R. K. and A. J. Miller, 1978: A study of the atmospheric energetics of a six-layer operational forecast model. Mon. Wea. Rev., 106, 607-613.

Miller, A. J., W. Collins and D. Dubofsky, 1975: The NMC operational global energy program. Office Note 109, National Meteorological Center, Washington, D.C., 13 pp.

Phillips, N. A., K. Campana, and M. Mathur, 1977: Data and analysis errors on 9 January 1977. Office Note 160, National Meteorological Center, Washington, D.C., 24 pp.

Shuman, F., and J. Hovermale, 1966: An operational six-layer primitive equation forecast model. J. Appl. Meteor. 7, 525-547.

Shuman, F. G., and J. D. Stackpole, 1969: The currently operational NMC model, and results of a recent simple numerical experiment. Proc. of the WMO/IVGG Symposium on Numerical Weather Prediction, Tokyo, Nov. 26 - Dec. 4, 1968. JMA, Tokyo, II-85 to II-98.

Stackpole, J. D., 1978: The NMC 9-layer global primitive equation model on a latitude-longitude grid. Office Note 178, National Meteorological Center, Washington, D.C., 31 pp.

Appendix I

The 3(xl) Model Evaluation

Instructions to FD forecasts jury.

Attached is a set of four forecasts: The operational 6LPE and three different contenders for possible replacement of the PE. The 6LPE model forecasts are identified as such on the maps but the others are not; they are only identified as models A, B, and C. All forecasts are from the same initial conditions. The verifying analyses and 500 mb height error charts for the three forecasts periods (24, 48, & 84 hours) are also attached.

Note: For any one initial date the model identified as "A", for example, will be the same throughout the forecast from that date. However, for a different initial date the "A identifier may mean a different model - we are randomizing this to avoid possibilities of bias.

In this evaluation we are asking you to rank the four with respect to a number of rather specific portions of the forecasts. They are generally those things for which there are operational requirements that the LFM model doesn't or cannot satisfy. There are five such areas of specific interest.

I. In support of Marine forecasting (and the contention by some that the LFM is not as useful as one might like in the western U.S. and Alaska), please evaluate the sea level pressure and 500 mb - 1000 mb thickness progs over the Western Atlantic and (combined) Eastern Pacific Alaska and Western Region areas. These evaluations are for 24 and 84 hour forecasts only.

II. As an estimate of ATA wind forecast quality, consider the location and strength of the 300 mb jet over the Pacific North America, and the Atlantic for the 24 hour forecast.

III. Please consider how the precipitation forecasts look over eastern and western U.S. at 24 and 48 hour (East and West of 105°), both in areal coverage and quantitative amounts.

IV. In support of the extended forecast program please evaluate the 84 hour forecasts of upper air, sea level and precipitation patterns (i.e. everything) but over the U.S. only.

V. Please address yourself to the general quality, consistency and credibility of the forecasts with particular reference to the 500 mb height and vorticity patterns. Are those maps useful in identifying and following short waves? Are they too smooth? Too noisy? If they were sent out to the field would we get back a lot of complaints, or praise?

A fill-in-the-blank answer sheet is attached.

Comments are most welcome. Use the space on the answer sheet and key your comments with the item number.

Have Fun!

3 (x1) Model Evaluation
Relative Evaluation of Four Models
(6L PE and Three Contenders)

Instructions: Indicate by the letter identifier (A, B, C, PE) which ranks 1st (best) 2nd, 3nd, & 4th (worst) for each area of interest and forecast hour. X's indicate positions for which no answer is called for.

Initial time of forecast ____________________

Evaluated by ____________________

Area Interest	Item	Forecast Hour and Rank 24				48				84			
		1	2	3	4	1	2	3	4	1	2	3	4
I. Marine Alaska & West Sea Level & Thickness													
Western Atlantic	1									X	X	X	X
East Pacific, Alaska & West	2									X	X	X	X
II. ATA Location & Strength of 300mb Jet													
Atlantic	3	X	X	X	X					X	X	X	X
North America	4	X	X	X	X					X	X	X	X
Pacific	5	X	X	X	X					X	X	X	X
III. Precipitation													
Eastern U.S. coverage	6 a									X	X	X	X
Eastern U.S. amounts	6 b									X	X	X	X
Western U.S. coverage	7 a									X	X	X	X
Western U.S. amounts	7 b									X	X	X	X
IV. Extended Forecast													
U.S. Upper Air	8	X	X	X	X	X	X	X	X				
U.S. Sea Level	9	X	X	X	X	X	X	X	X				
East U.S. Precip coverage	10 a	X	X	X	X	X	X	X	X				
East U.S. Precip amounts	10 b	X	X	X	X	X	X	X	X				
West U.S. Precip coverage	11 a	X	X	X	X	X	X	X	X				
West U.S. Precip amounts	11 b	X	X	X	X	X	X	X	X				
V. Quality, Consistency, & Credibility (particularly 500mb)													
Hemispheric	12												
North America & Waters	13												

Comments (specify item number)

Summary: I like model A B C PE (circle one) best

4. Conclusions, Recommendations, and Spin-offs. As should be apparent by now, the HFM was selected as the candidate to replace the 6LPE as NMC's operational forecast model. However, it was equally apparent that introducing the HFM would be of no benefit to the stratospheric forecasts. The improvements shown by the NGM and 9LH models, on the other hand, suggested that if the HFM could be modified to increase its resolution and reach higher levels in the stratosphere, such a seven layer PE model (7LPE) would combine the best features of the HFM and the higher reaching NGM and 9LH models. Such a program of modifications was undertaken, as the addition of the 7th layer could be accomplished without any increase in execution times of the HFM model. The device was to replace the layer of constant potential temperature (the "thetasphere") that ostensibly served as a computational upper boundary layer for the HFM with a third stratospheric layer fully coupled to the remainder of the model below. The upper boundary conditions then became constant pressure (with a top set at 50 mb) and zero vertical velocity. The substitution of a meteorological layer for the "thetasphere" layer was economically feasible as computations for the latter already involved forecasts of u, v, and p, thus one additional forecast equation for θ was balanced off by deletion of one equation for p tendency.

We were sanguine over the prospect of removing the computational boundary layer from the model (it has been in place since the model's inception) by virtue of having performed similar surgery to the global forecast model some time ago, making it a 9 layer model. No deleterious effects were noted - indeed improvements took place. Further preimplementation testing of the winner of the four model run-off took place with the model in the form of the 7LPE. Those results are discussed below.

Other side effects of the test series came along - the rather poor performance of the NGM model in forecasting precipitation (far too large areas of rain forecast) have lead the NGM modelers to evaluations of how this came about and to remedial steps; analysis of one case in which the models were all rather poor and there were inconsistencies in their behavior has unearthed inadequacies in portions of NMC's initial data specifications (Phillips, et. al. 1977) with recommendations for improvement.

Other problems - the excessive smoothing in the 9LH model, the somewhat unsatisfactory energetics behavior of the NGM and HFM/7LPE, the poor precipitation forecasting in all the models - have served as spurs (of varying degrees of effectiveness) to the modelers most concerned.

III. Preimplementation Quality Assurance Tests

A. Execution of Tests

With the selection of the 7 layer fine mesh PE (7LPE) as the model of choice, it was, of course, appropriate to run another series of comparison tests: 7LPE vs 6LPE. This would serve both to assure all concerned that the 7LPE model was indeed an improvement over the 6LPE, and to give operational forecasters (NMC's forecasters, at least) a chance to become familiar with characteristics of the model prior to the operational implementation.

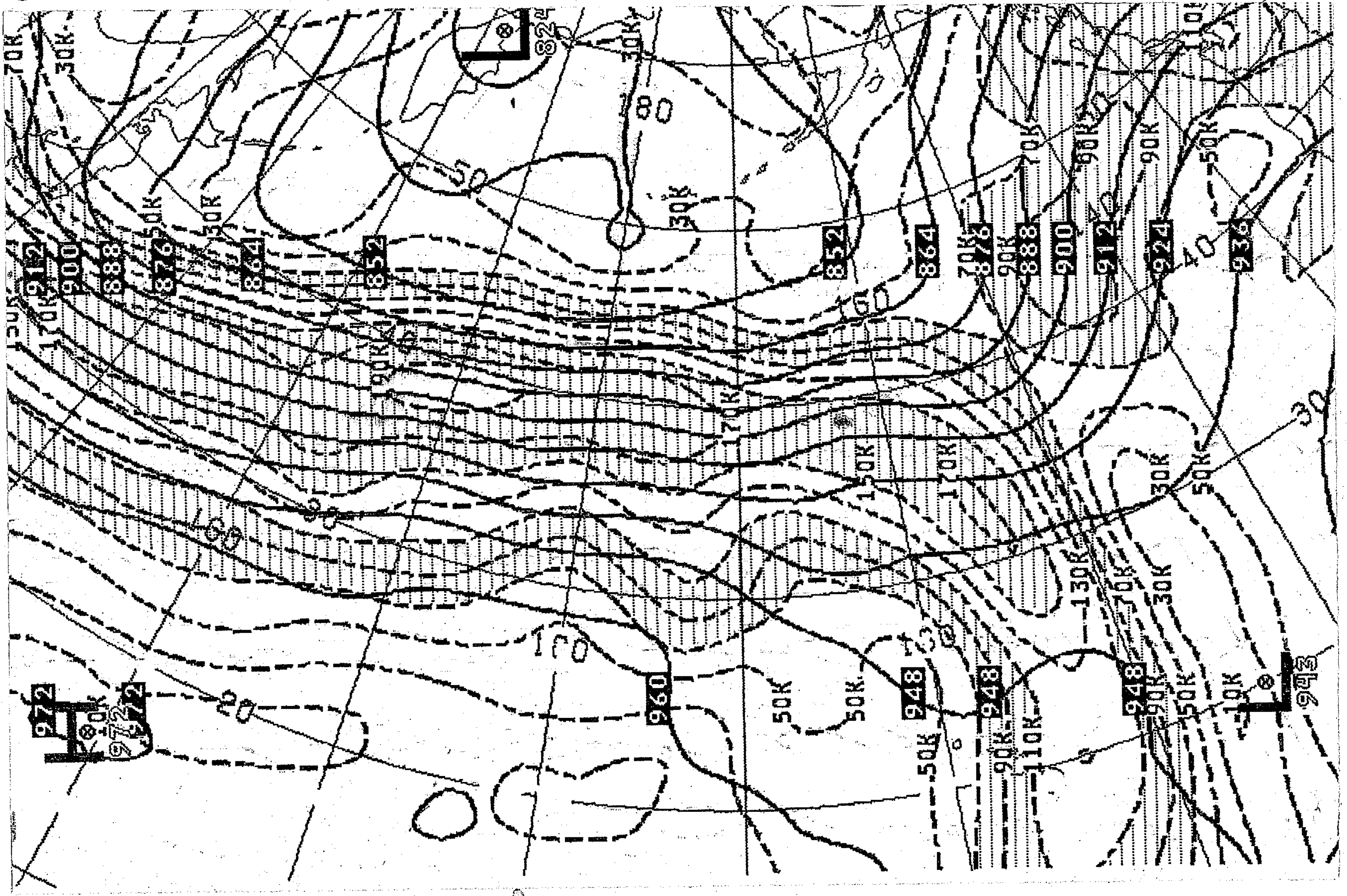


Fig. IV-1 - 36 hr 6LPE forecast 300 mb heights and Isotachs valid 12Z 18 FEB 77

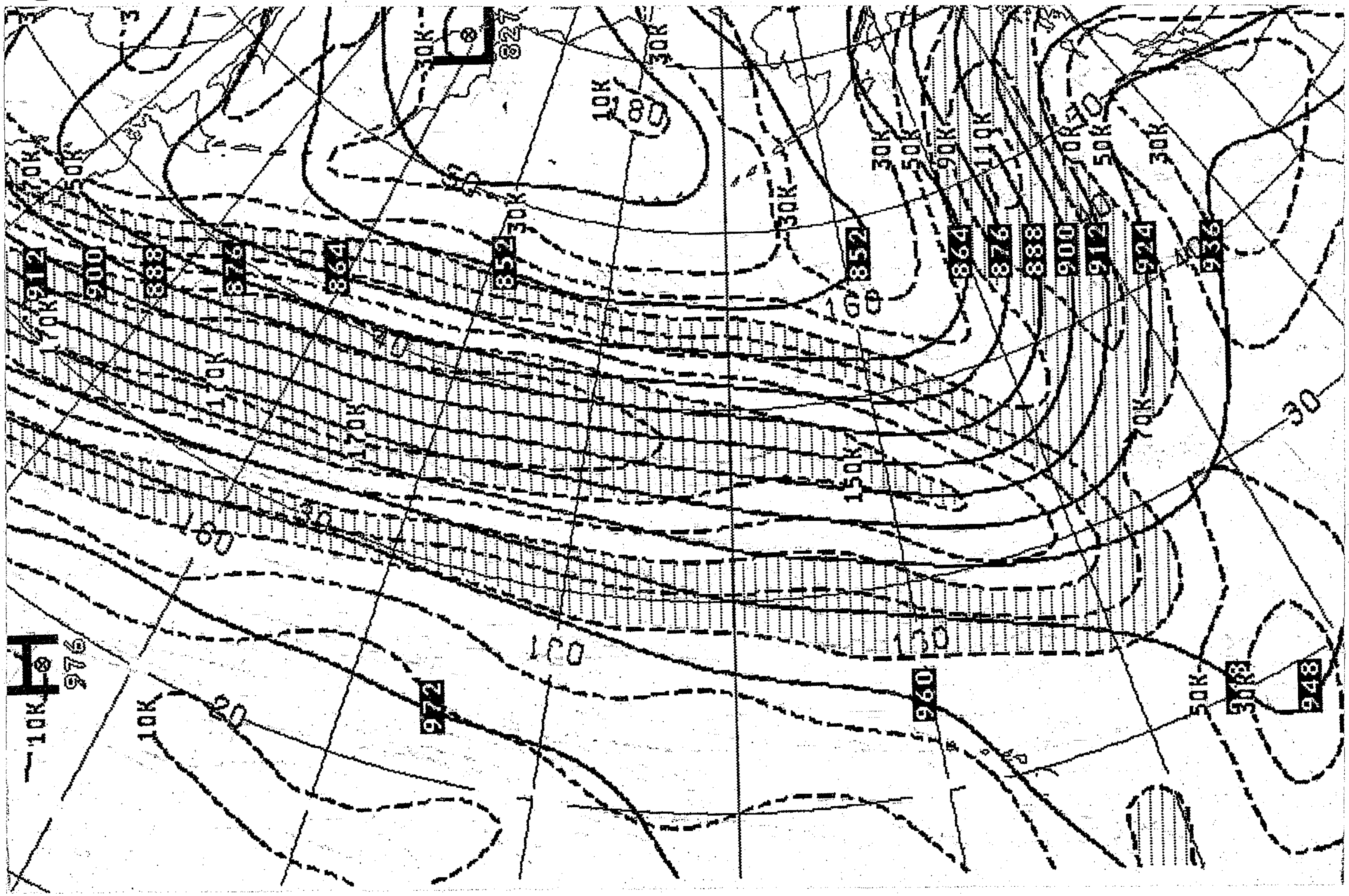


Fig. IV-2 - Analysis 300 mb heights and
Isotachs valid 12Z 18 FEB 77

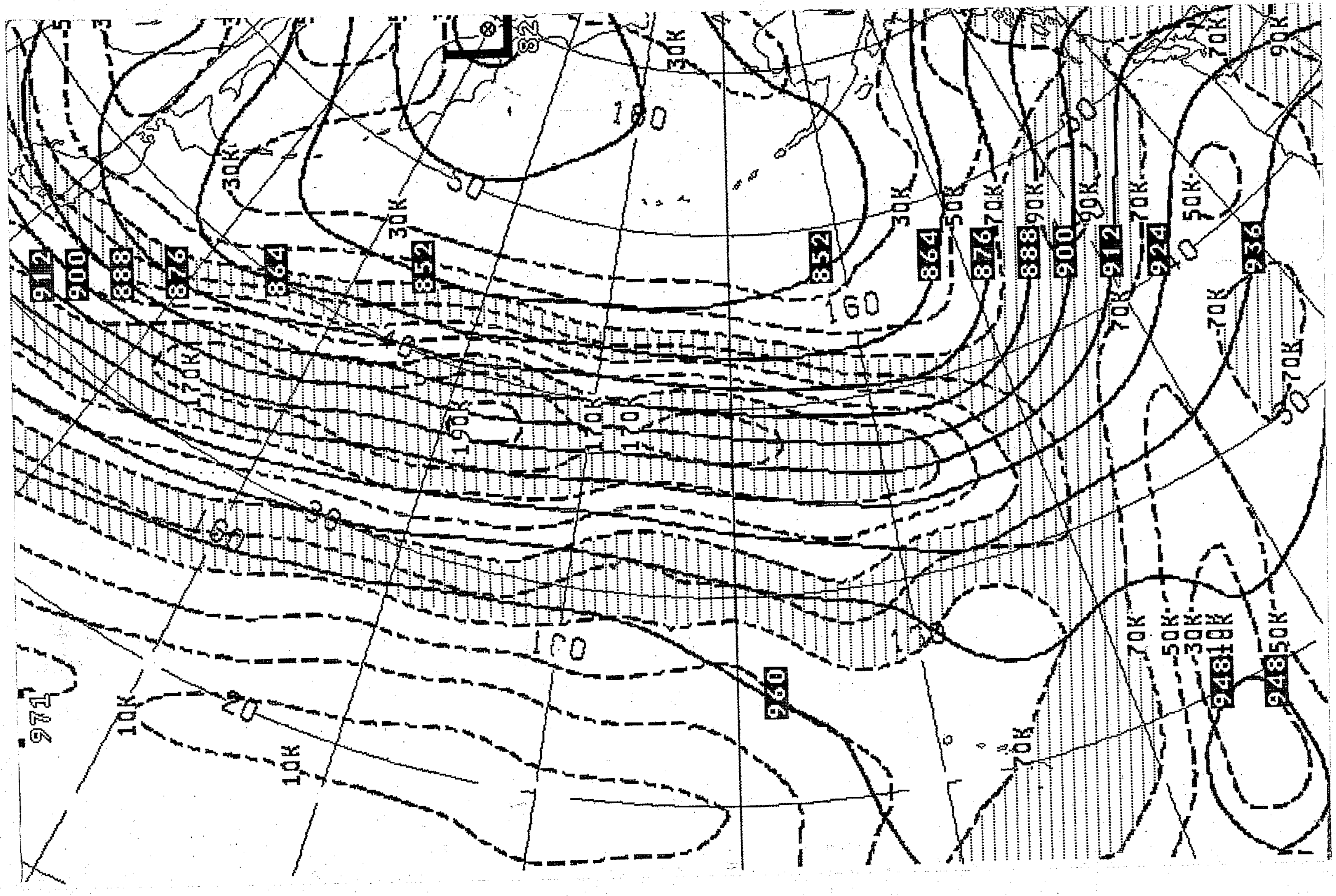
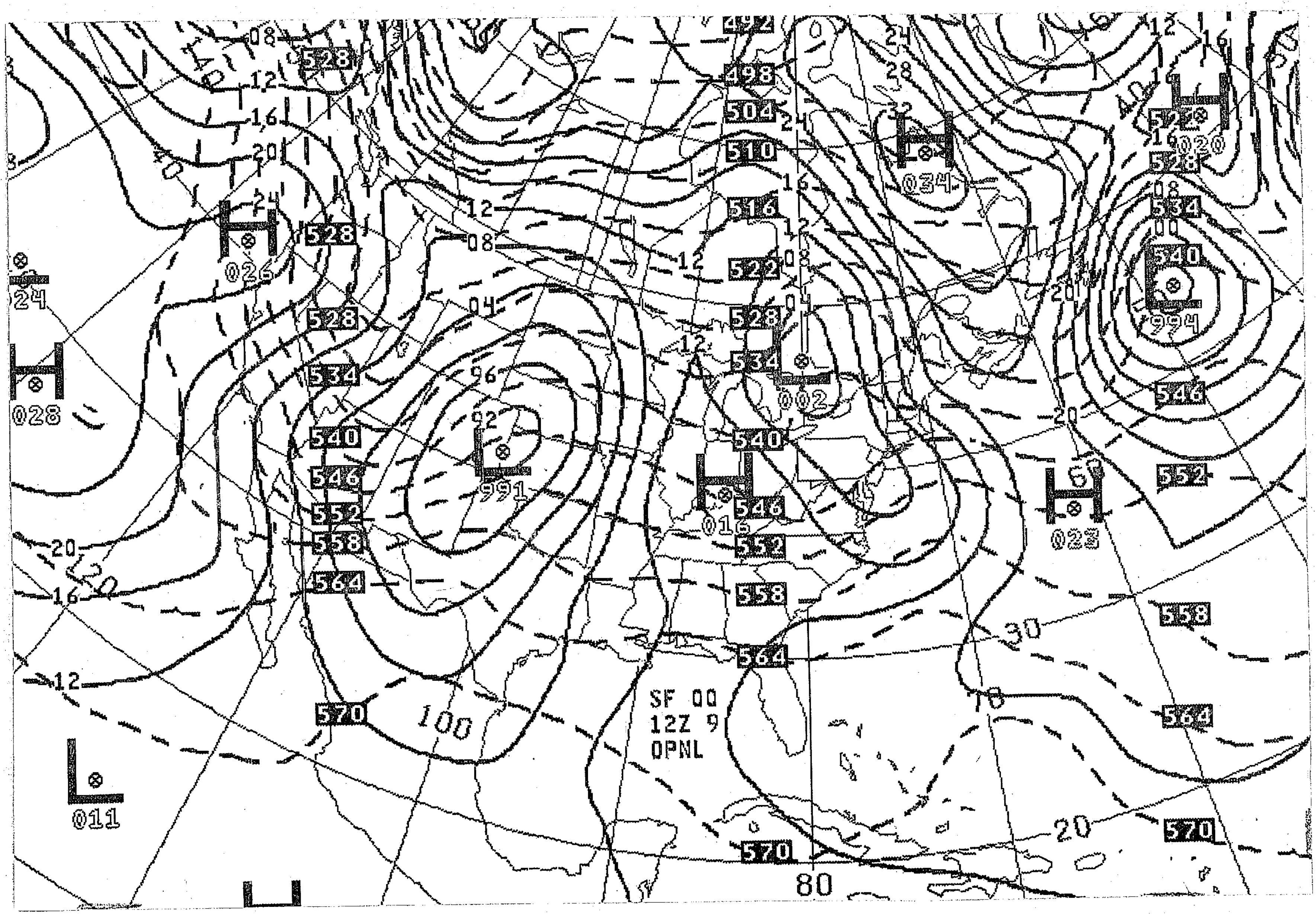
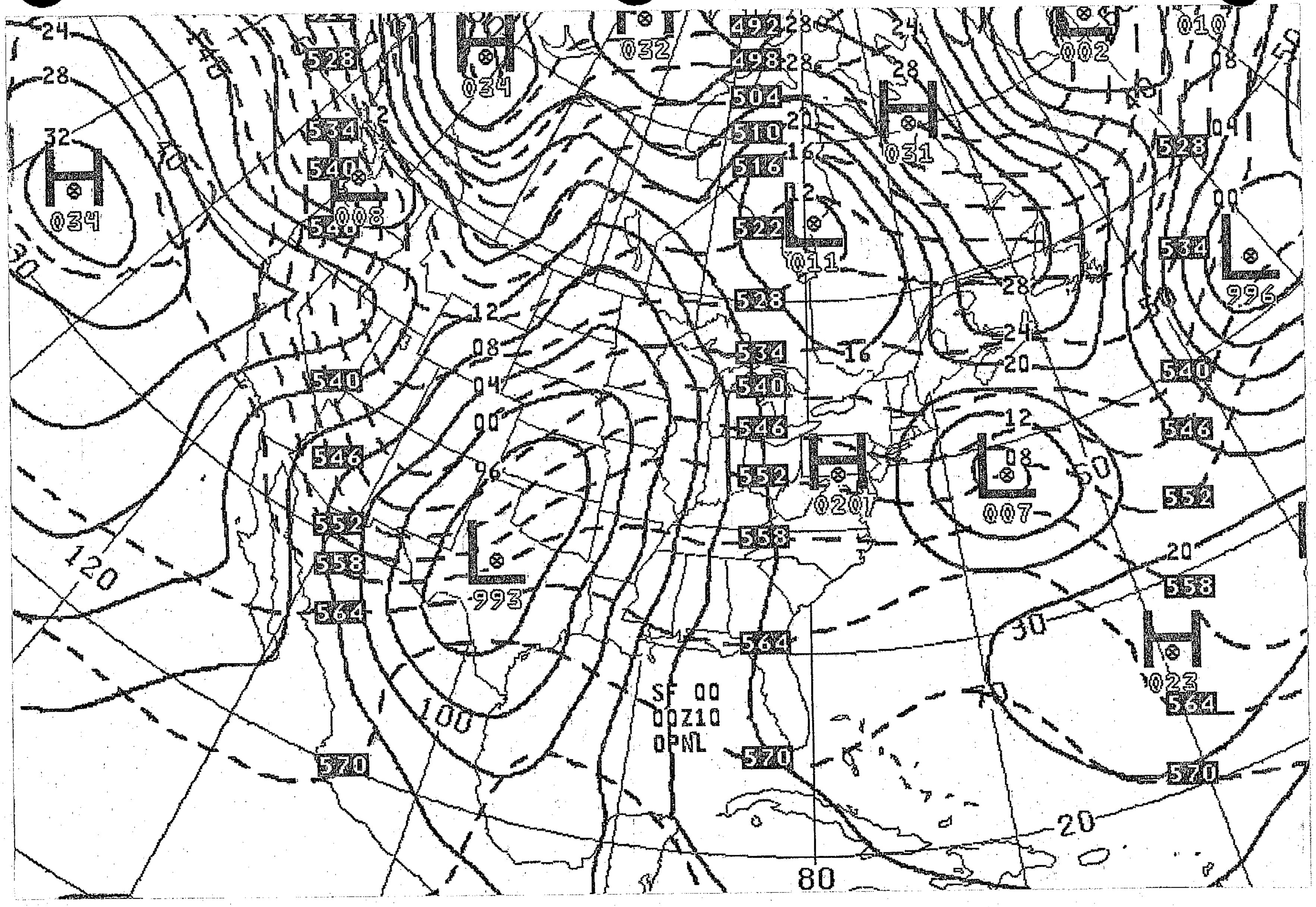


Fig. IV-3 - 36 hr 7LPE forecast 300 mb heights and Isotachs valid 12Z 18 FEB 77

Fig. IV-4 - Anal[illegible] Sea Level Pressure
500-[illegible] Thickness
12Z 9 JAN 75 - 12Z 11 JAN 75

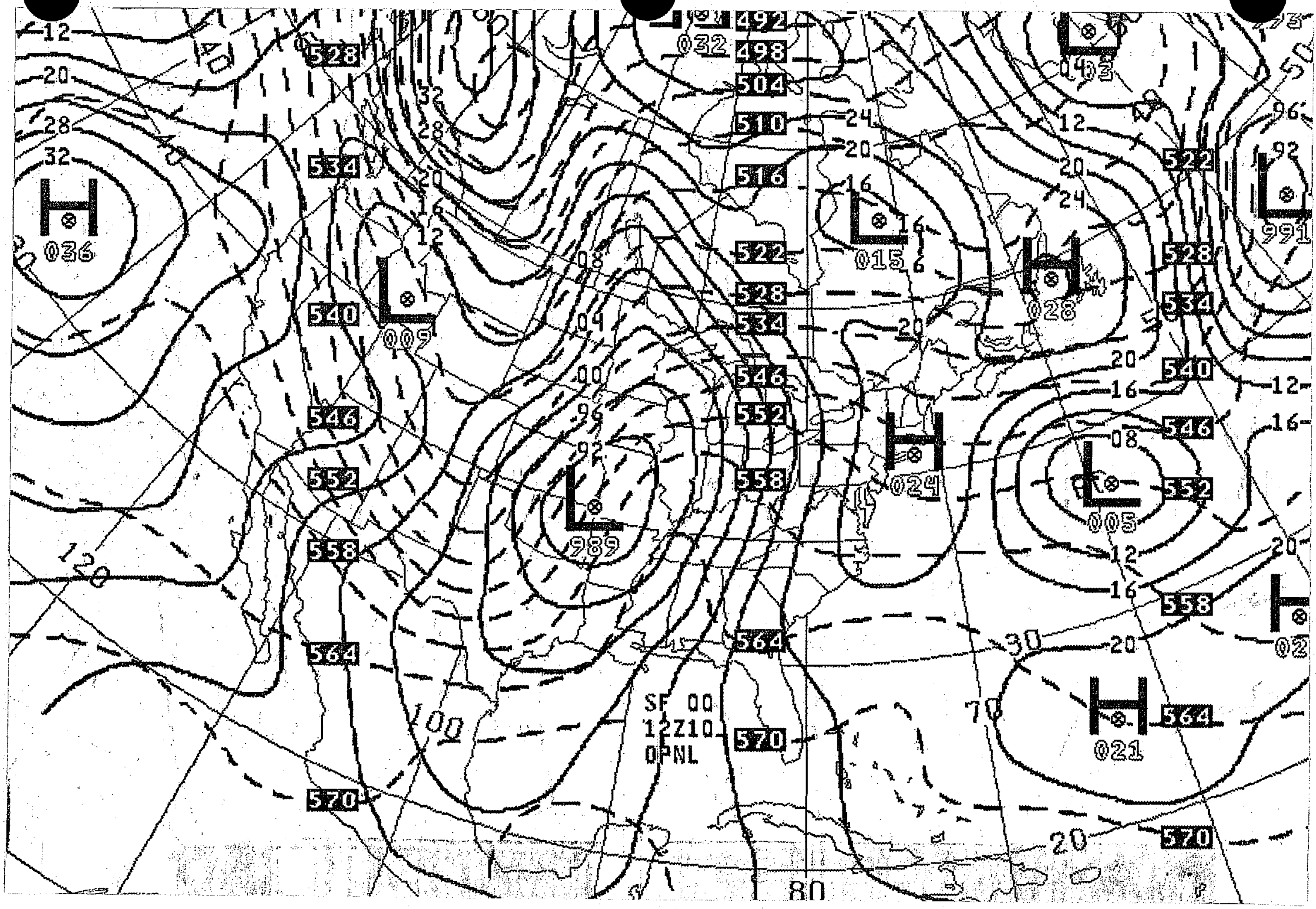


a
12Z 9 JAN 75



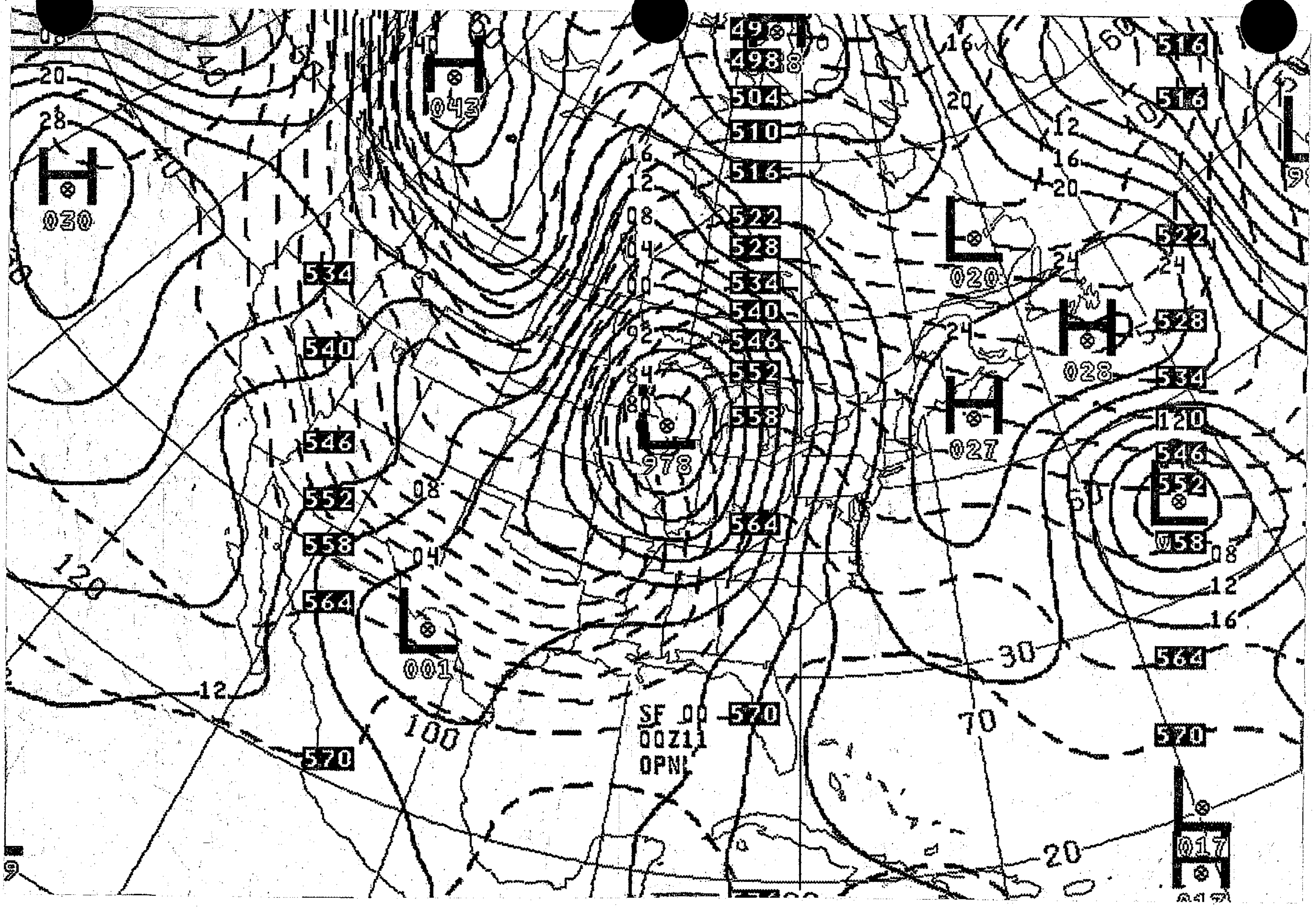
b

00Z 10 JAN 75



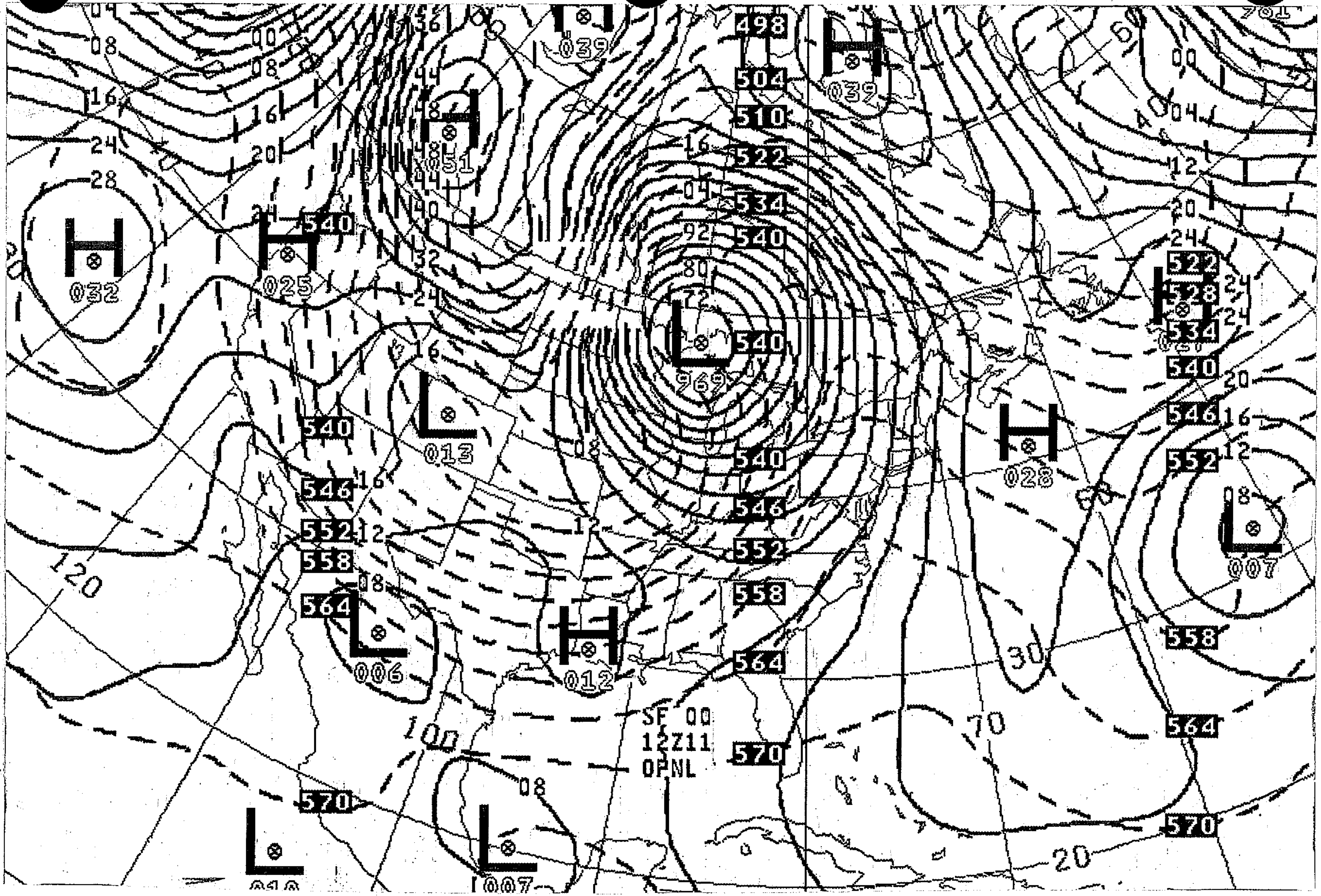
c

12Z 10 JAN 75



d

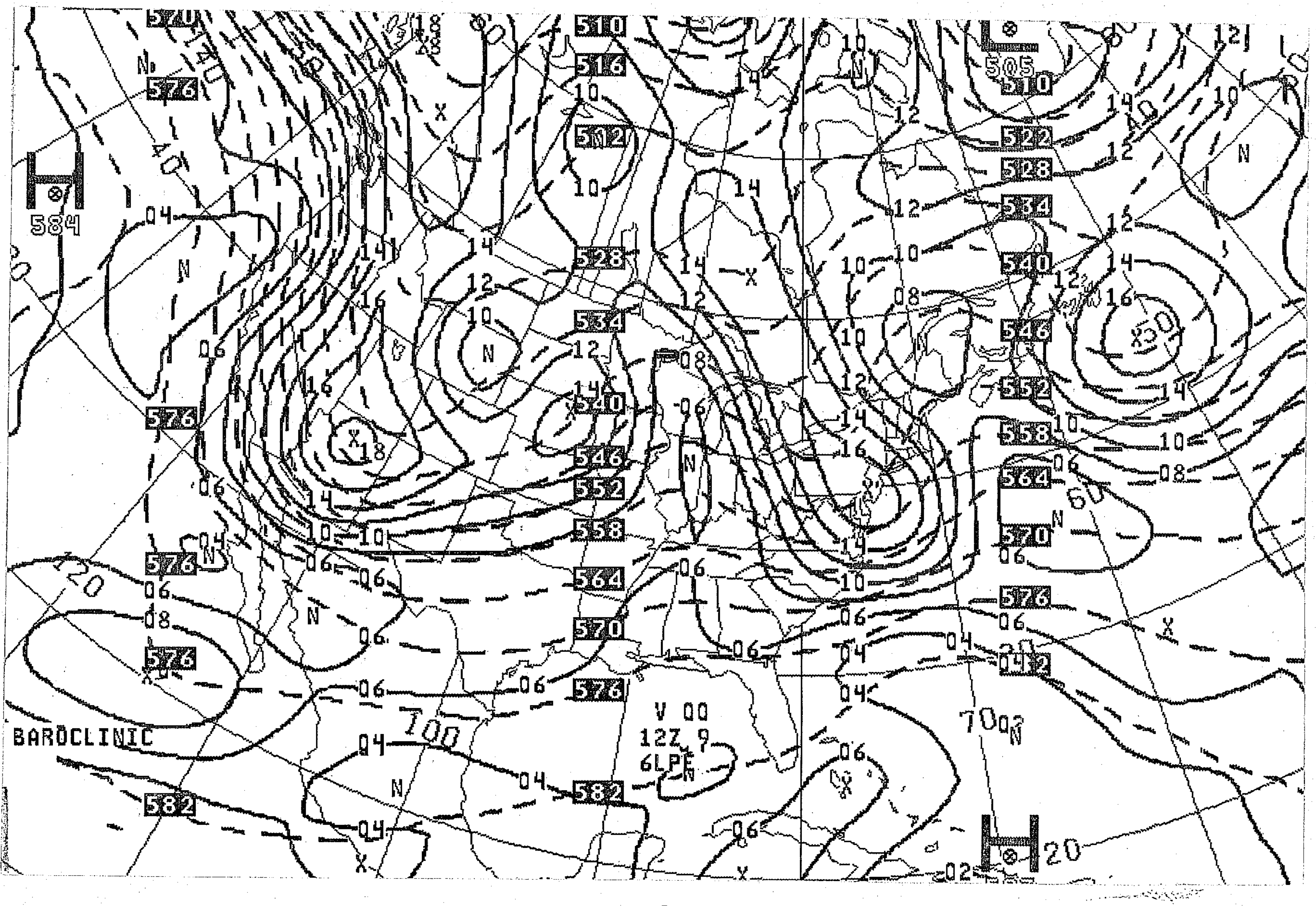
00Z 11 JAN 75



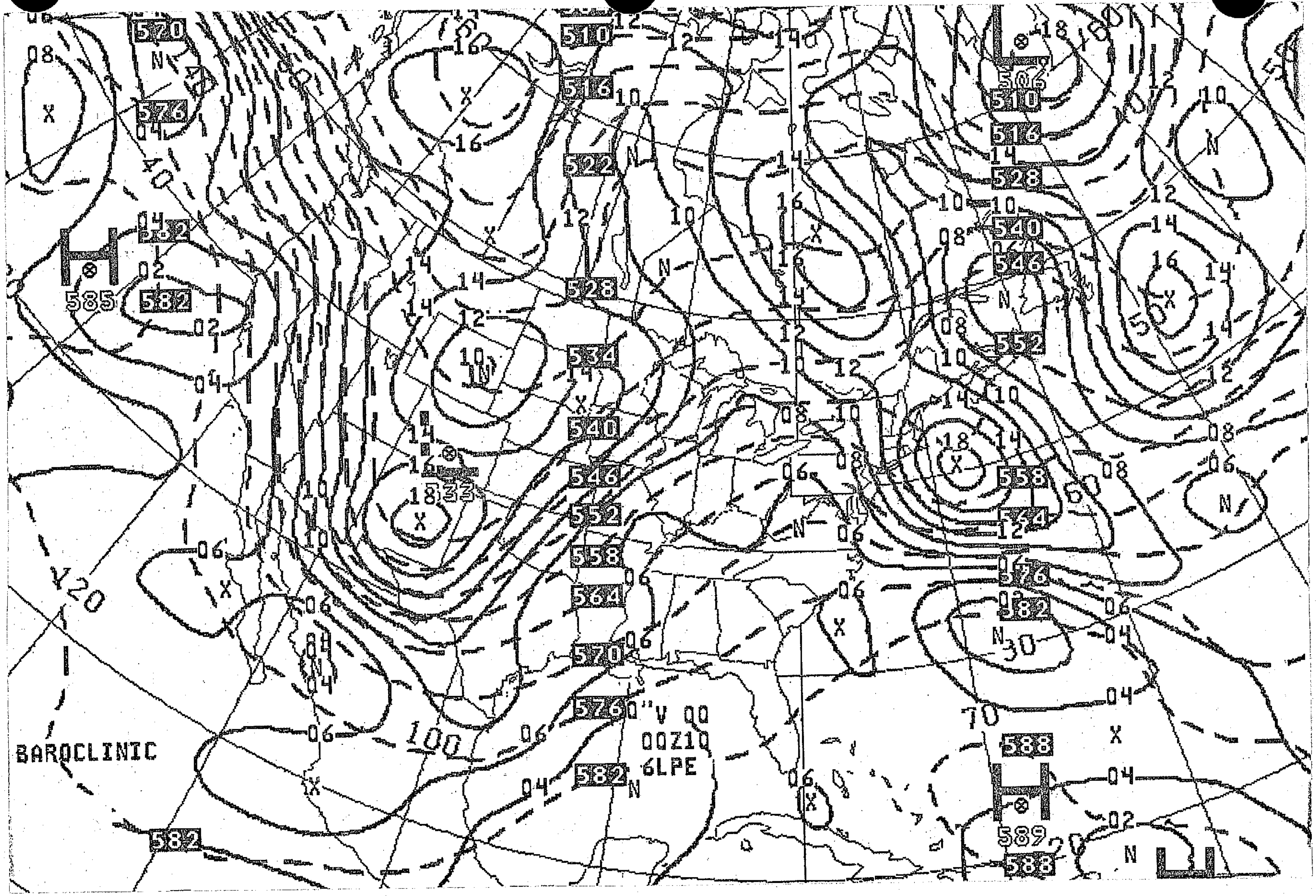
e

12Z 11 JAN 75

Fig. IV-5 - Analyses 500 mb height and vorticity

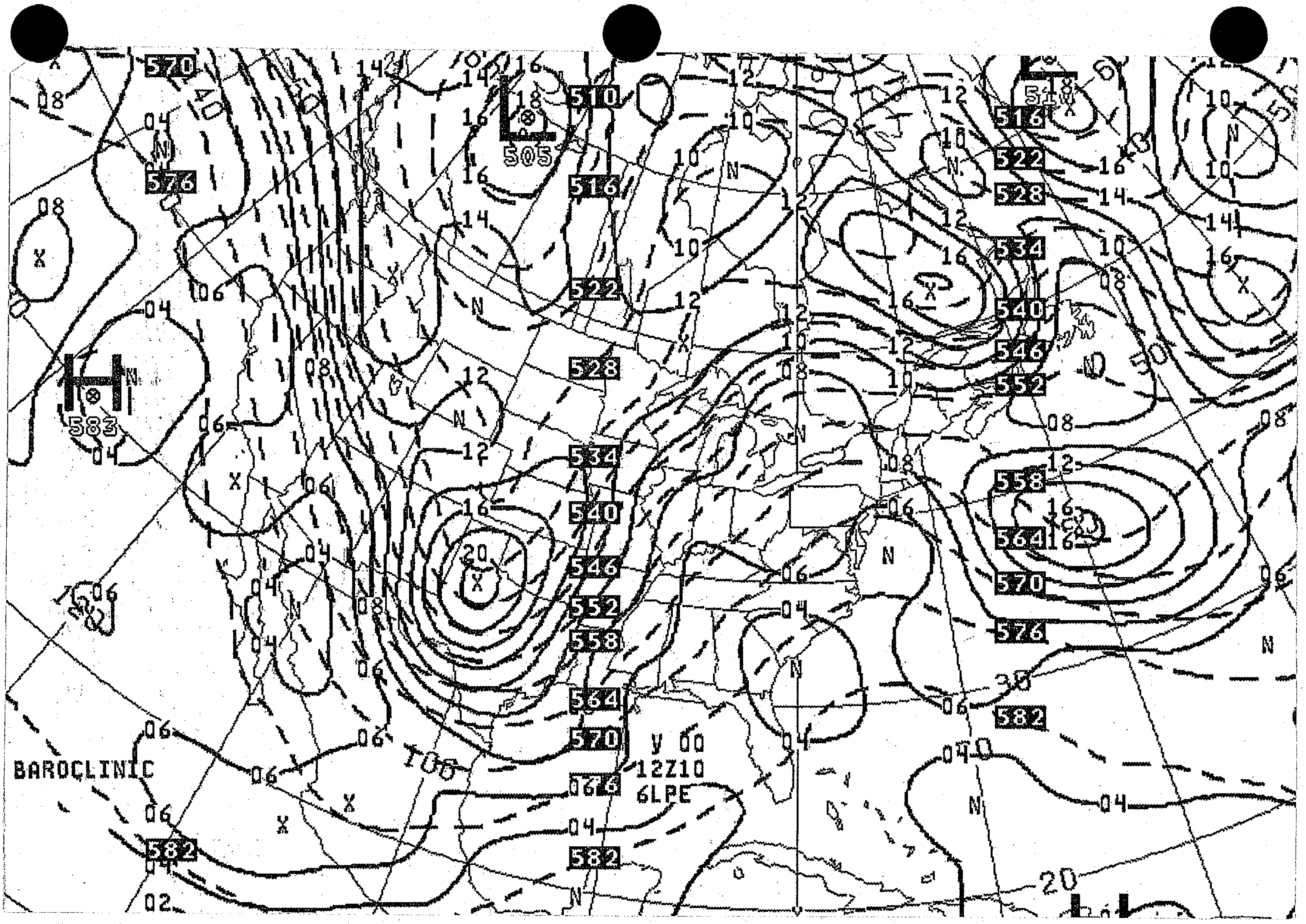


a
12Z 9 JAN 75



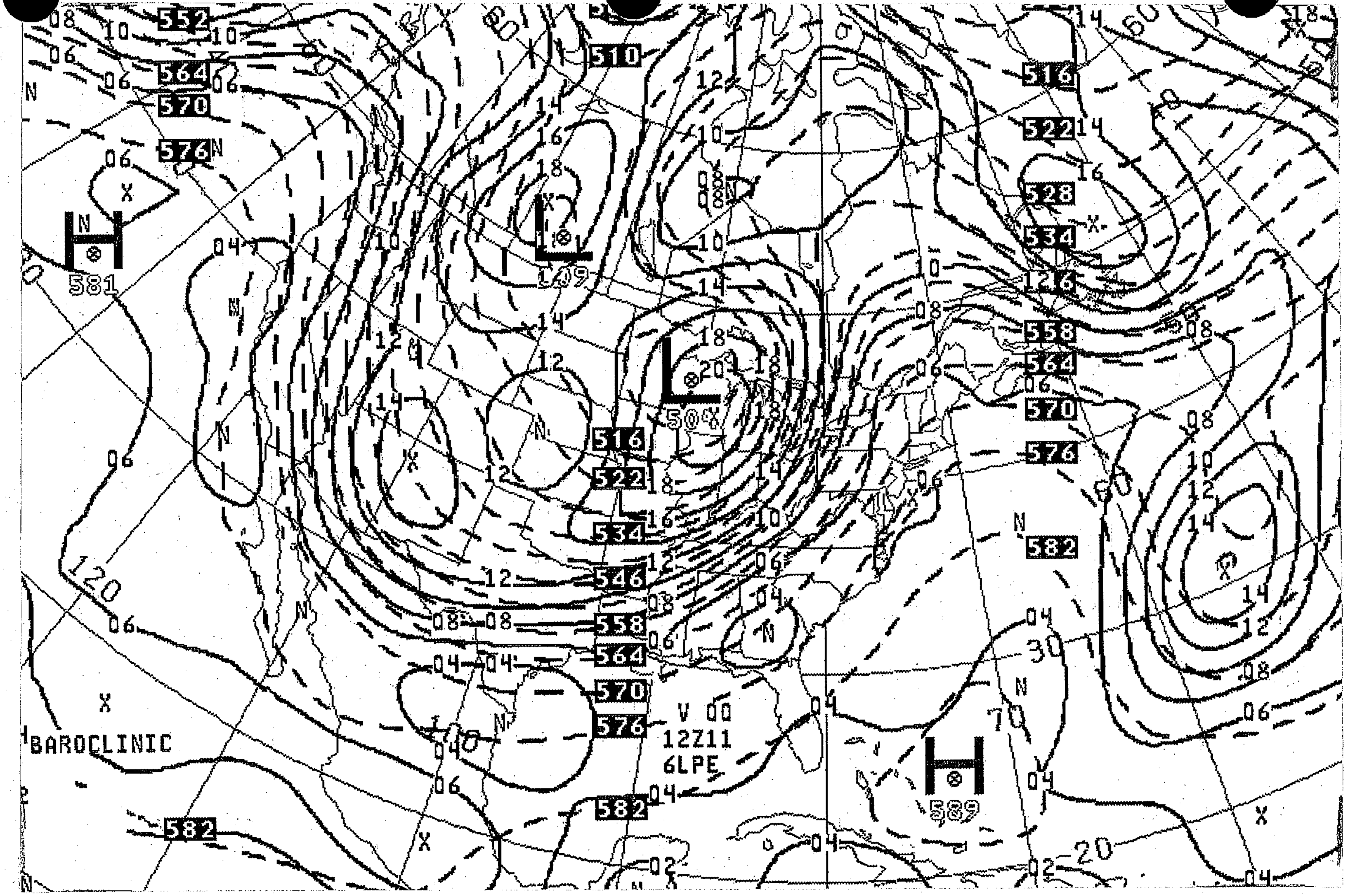
b

00Z 10 JAN 75



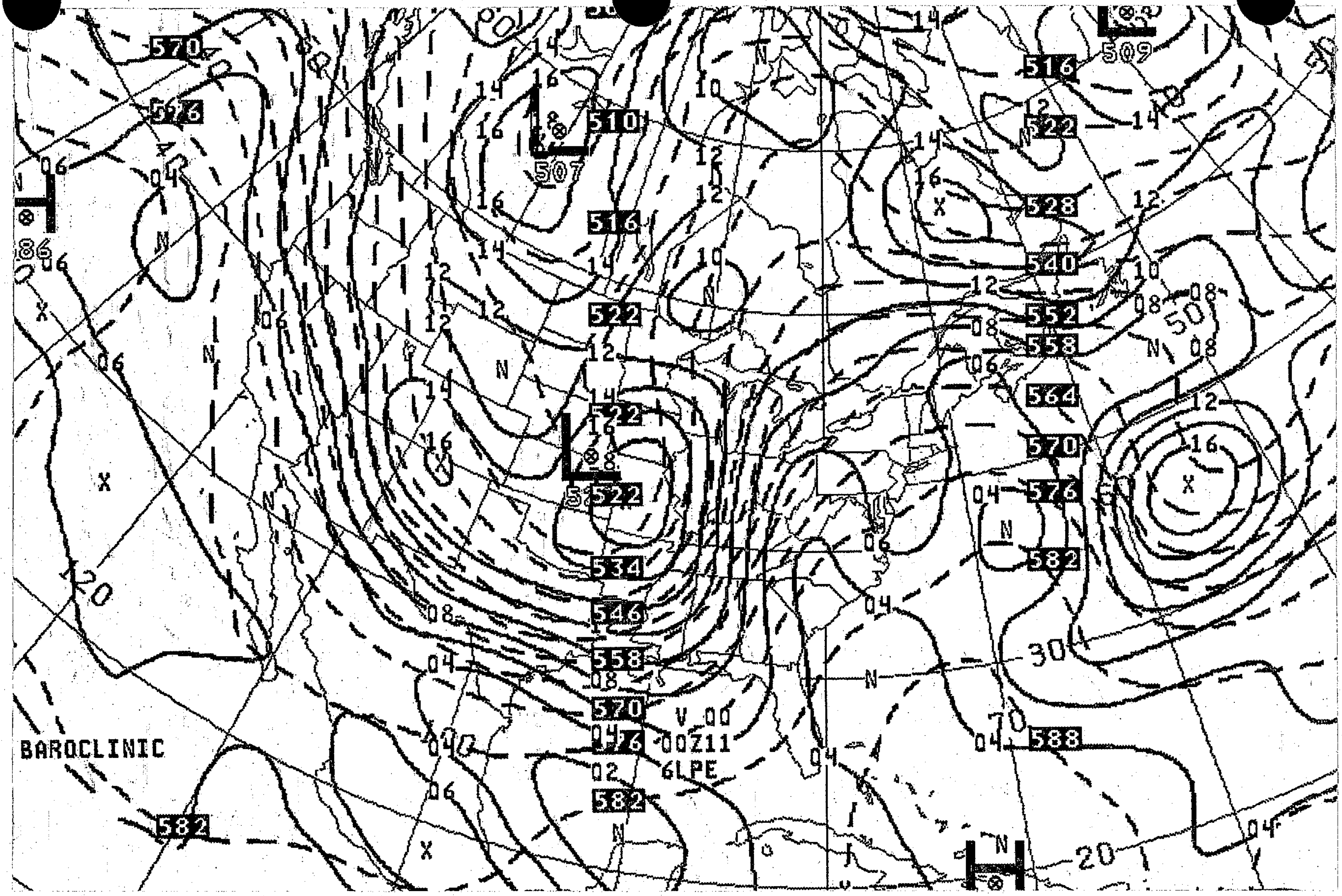
c

12Z 10 JAN 75



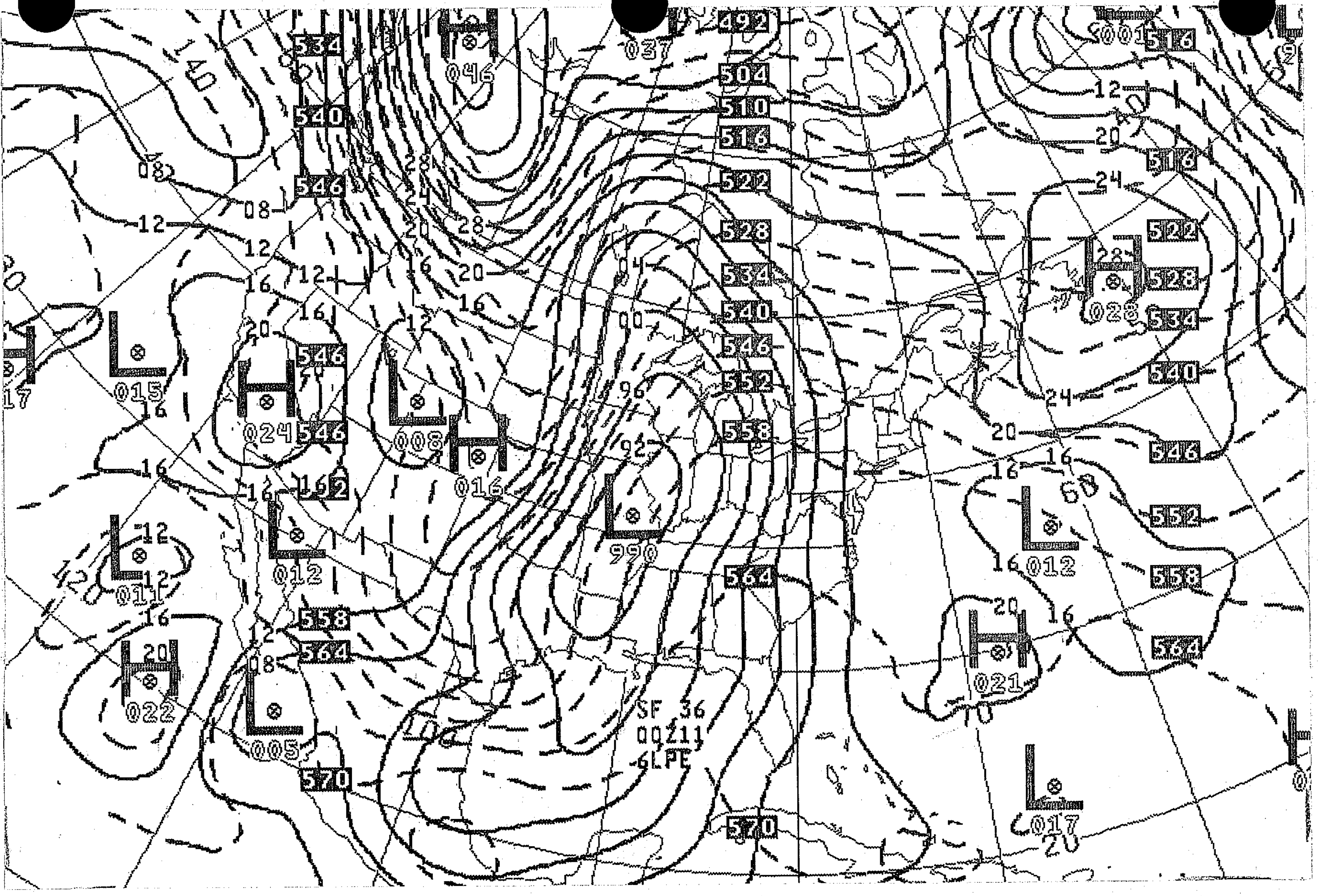
e

12Z 11 JAN 75



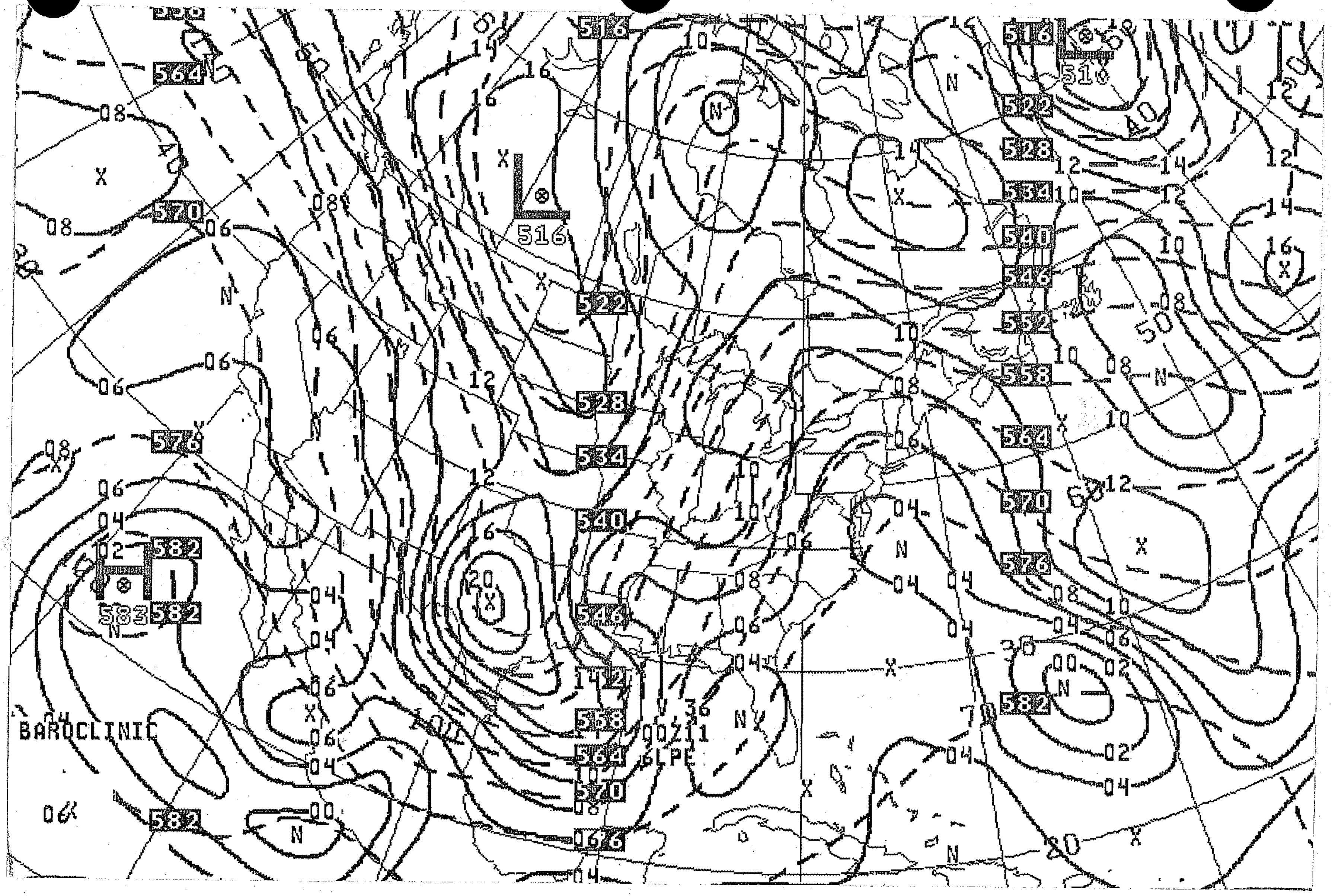
d

00Z 11 JAN 75



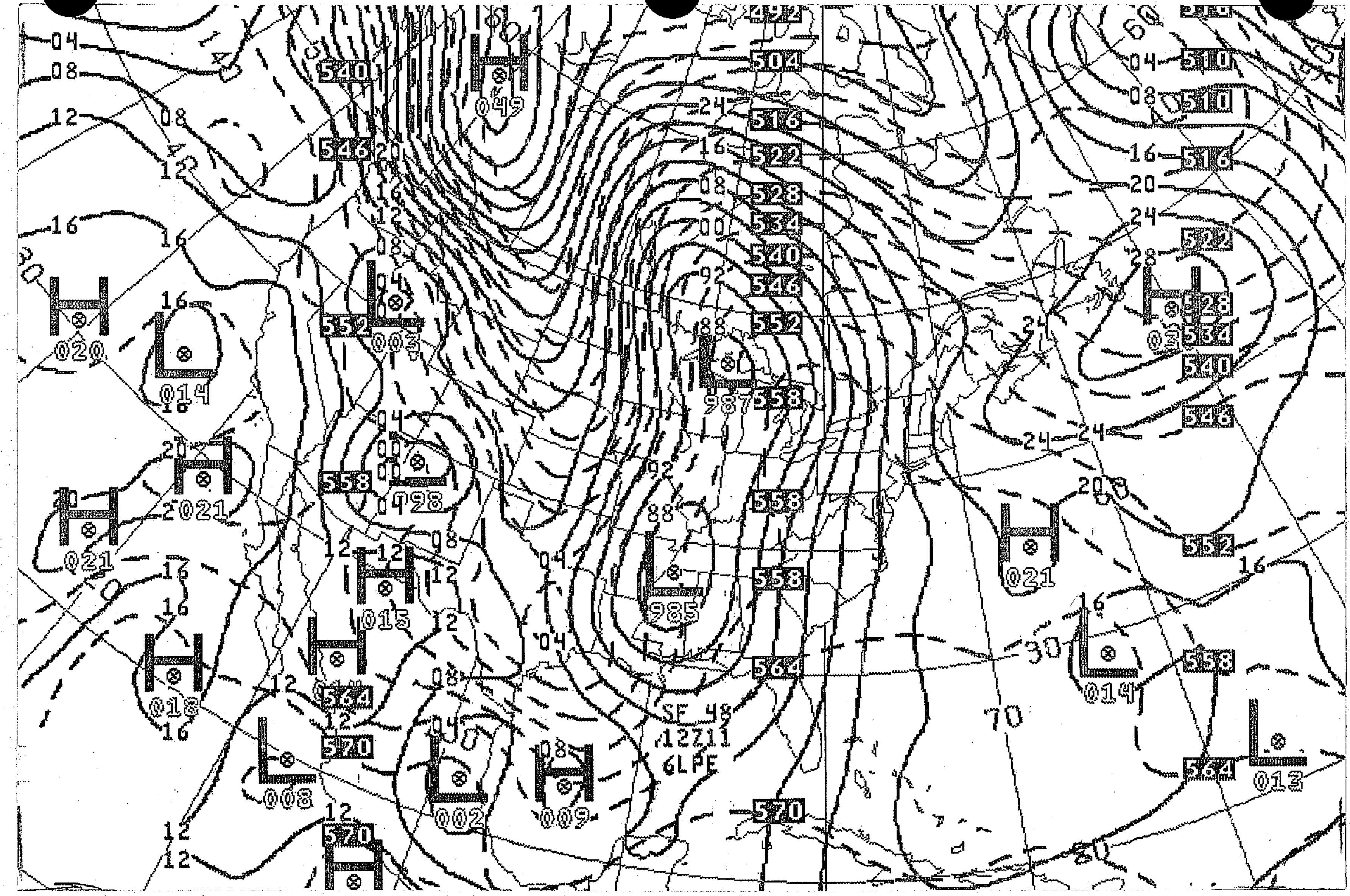
b

6LPE 36 hr forecast
valid 00Z 11 JAN 75



b

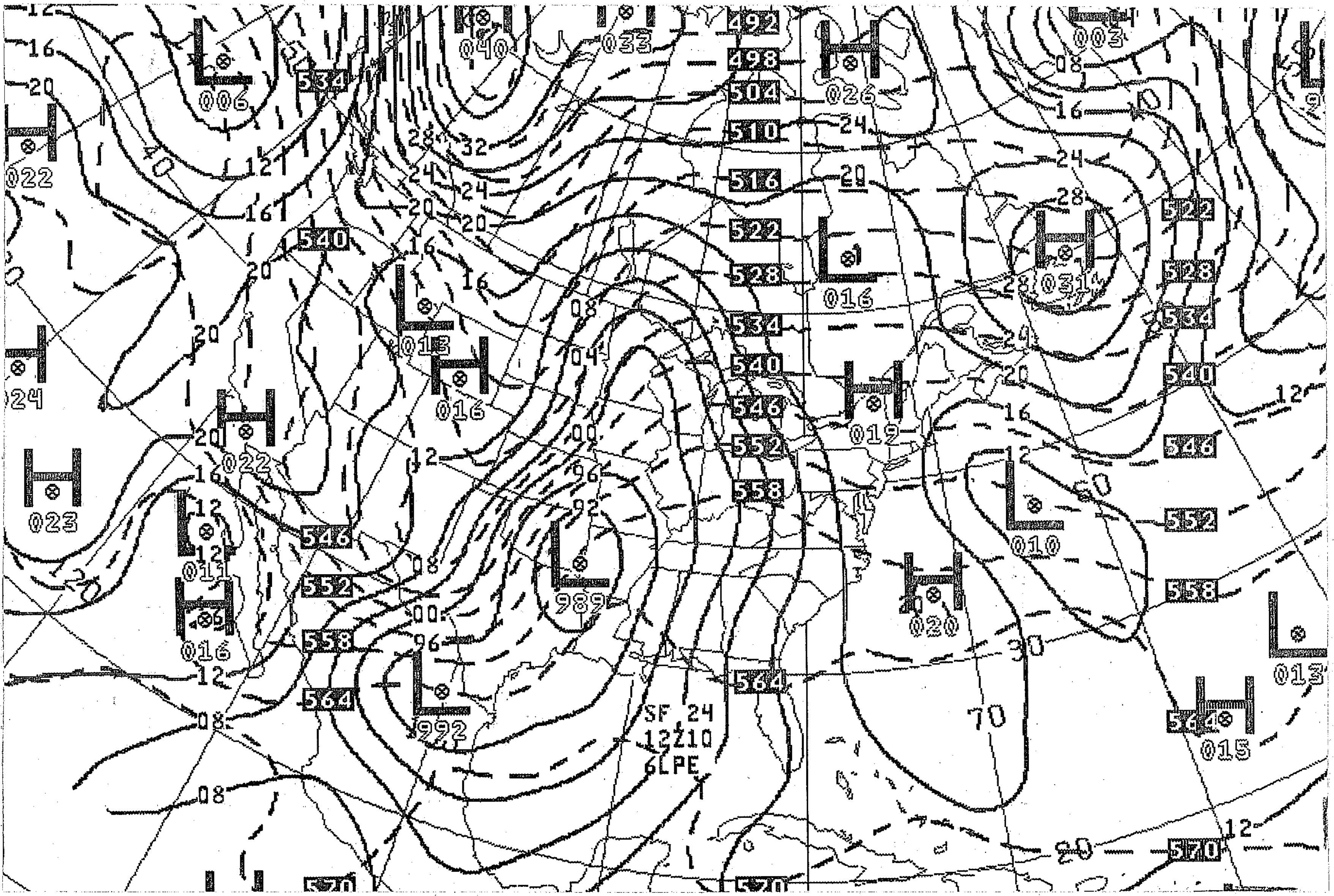
6LPE 36 hr Forecast
valid 00Z 11 JAN 75



c

6LPE 48 hr forecast
valid 12Z 11 JAN 75

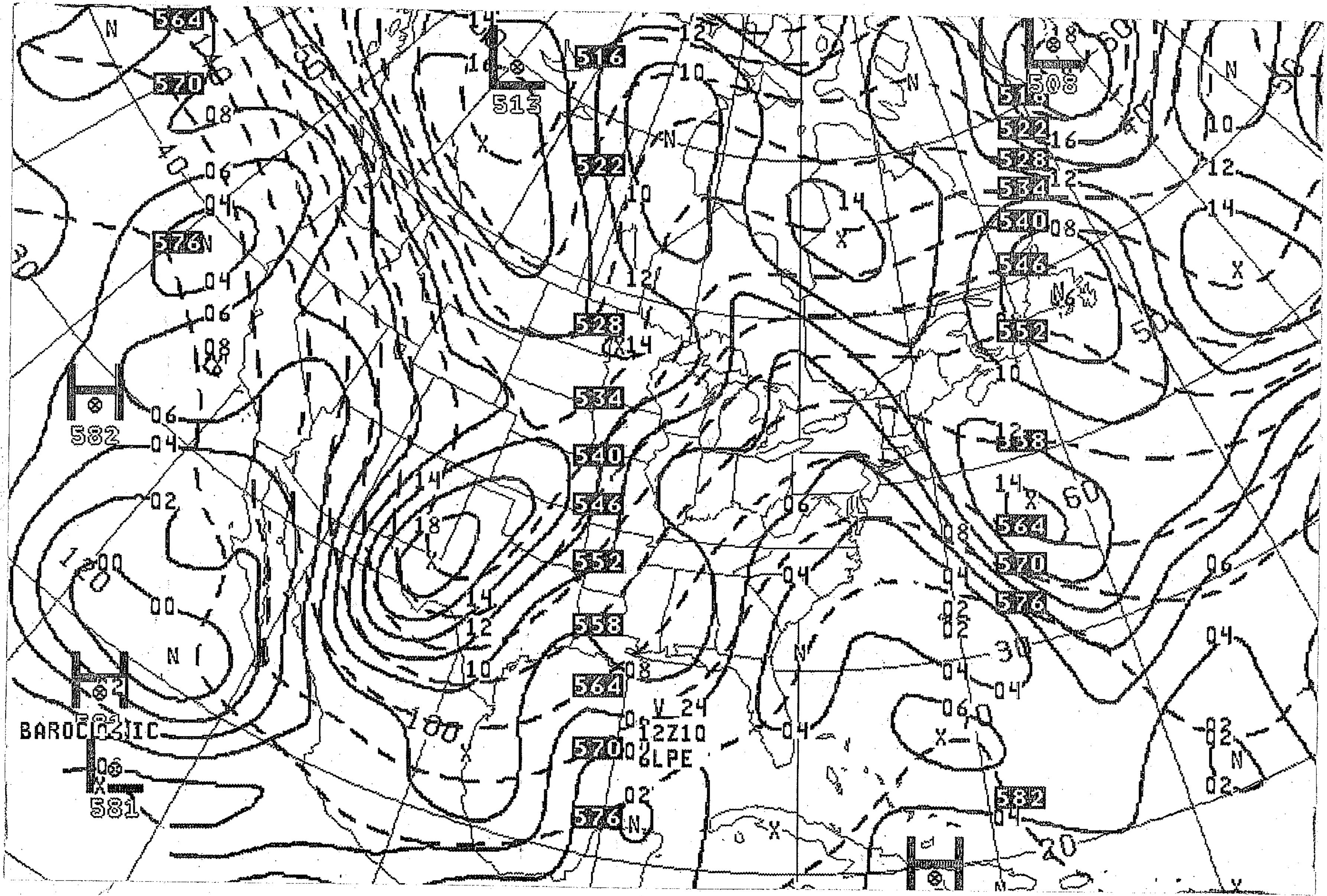
Fig. IV-6 - 6LPE 24, 36, 48 hr forecasts
Sea Level Pressure
500-1000 Thickness
Valid 12Z 10 JAN 75 - 12Z 11 JAN 75



a

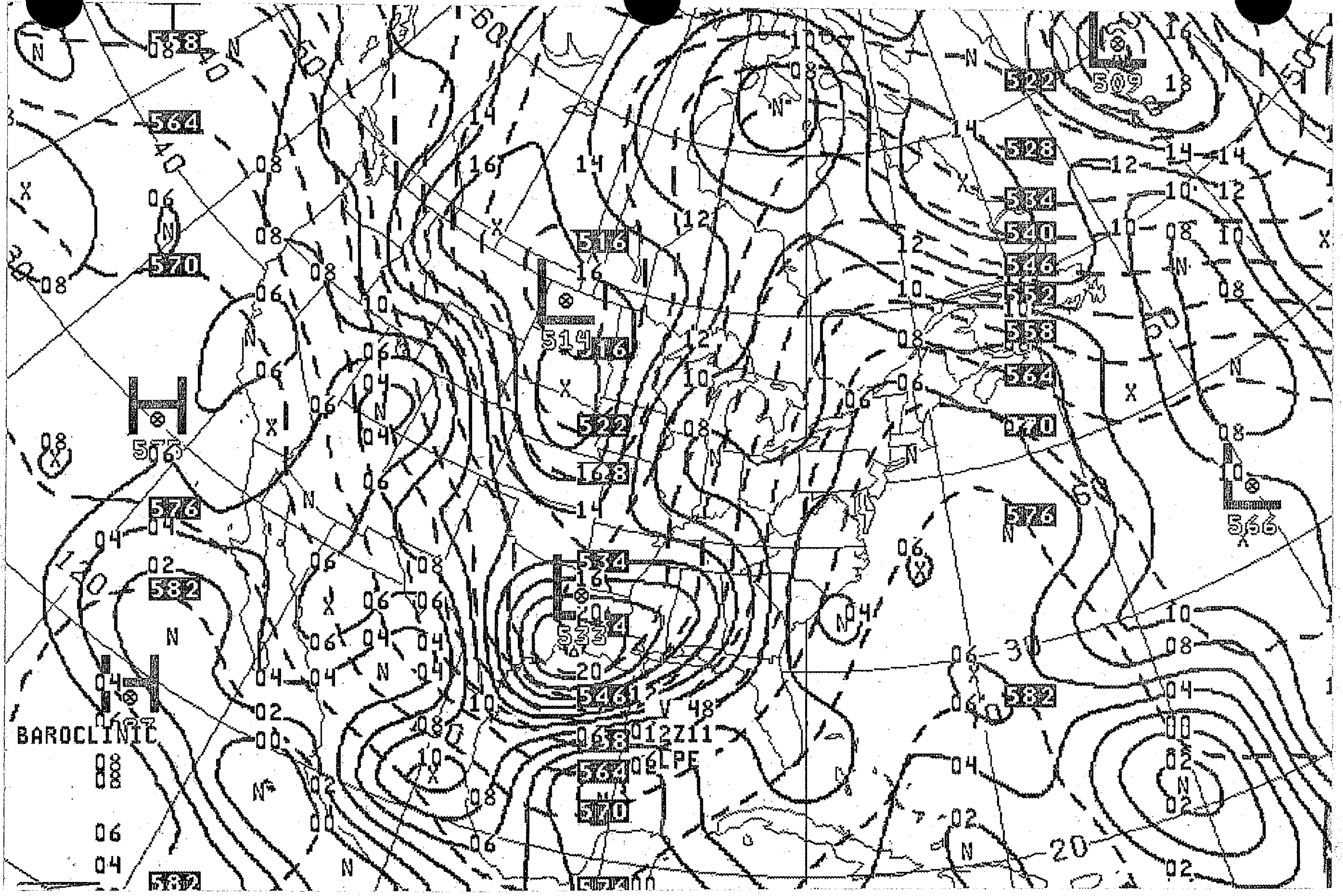
6LPE 24 hr forecast
valid 12Z 10 JAN 75

Fig. IV-7 - 6LPE 24, 36, 48 hr forecast
500 mb height and vorticity
valid 12Z 10 JAN 75 - 12Z 11 JAN 75



a

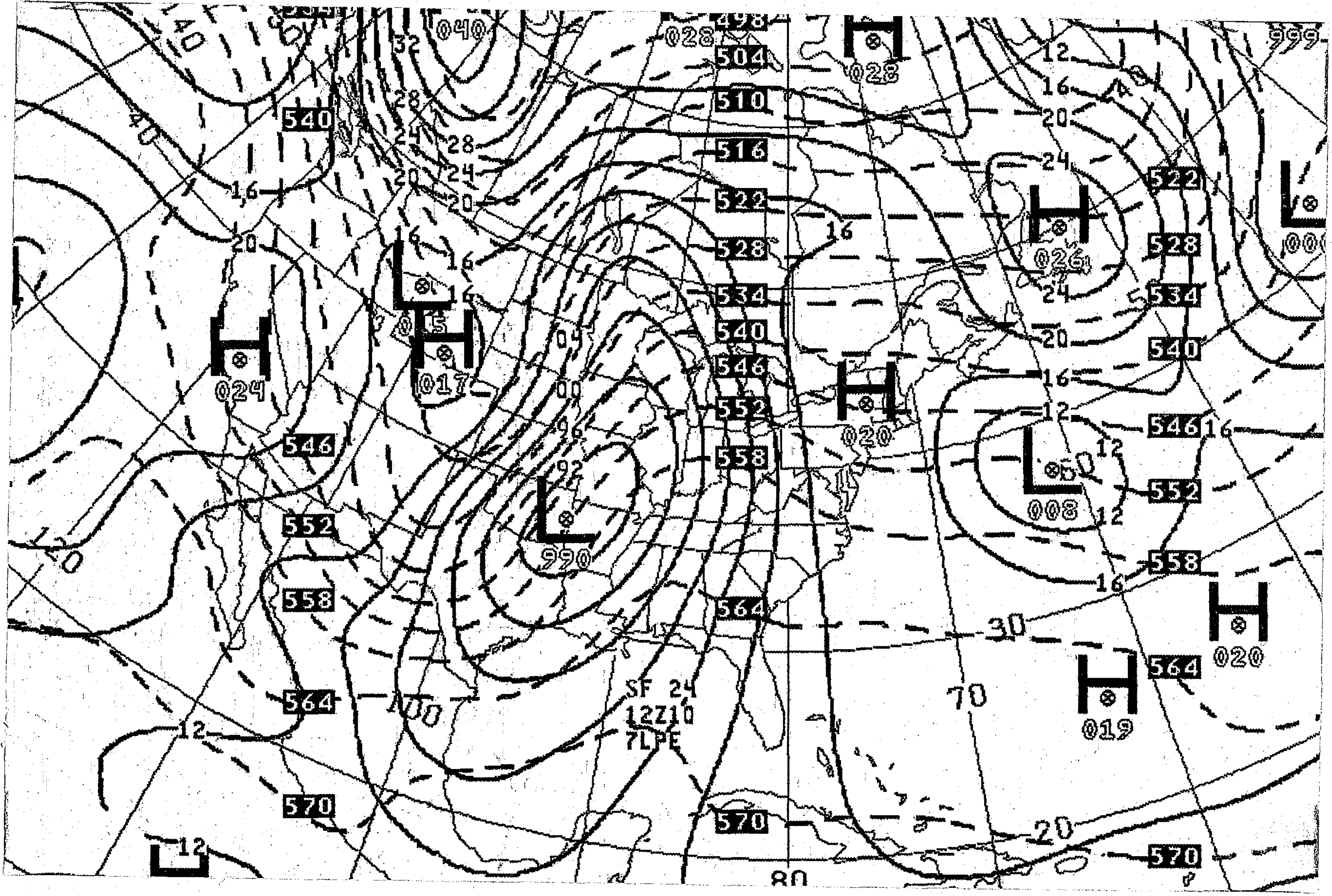
6LPE 24 hr forecast
valid 12Z 10 JAN 75



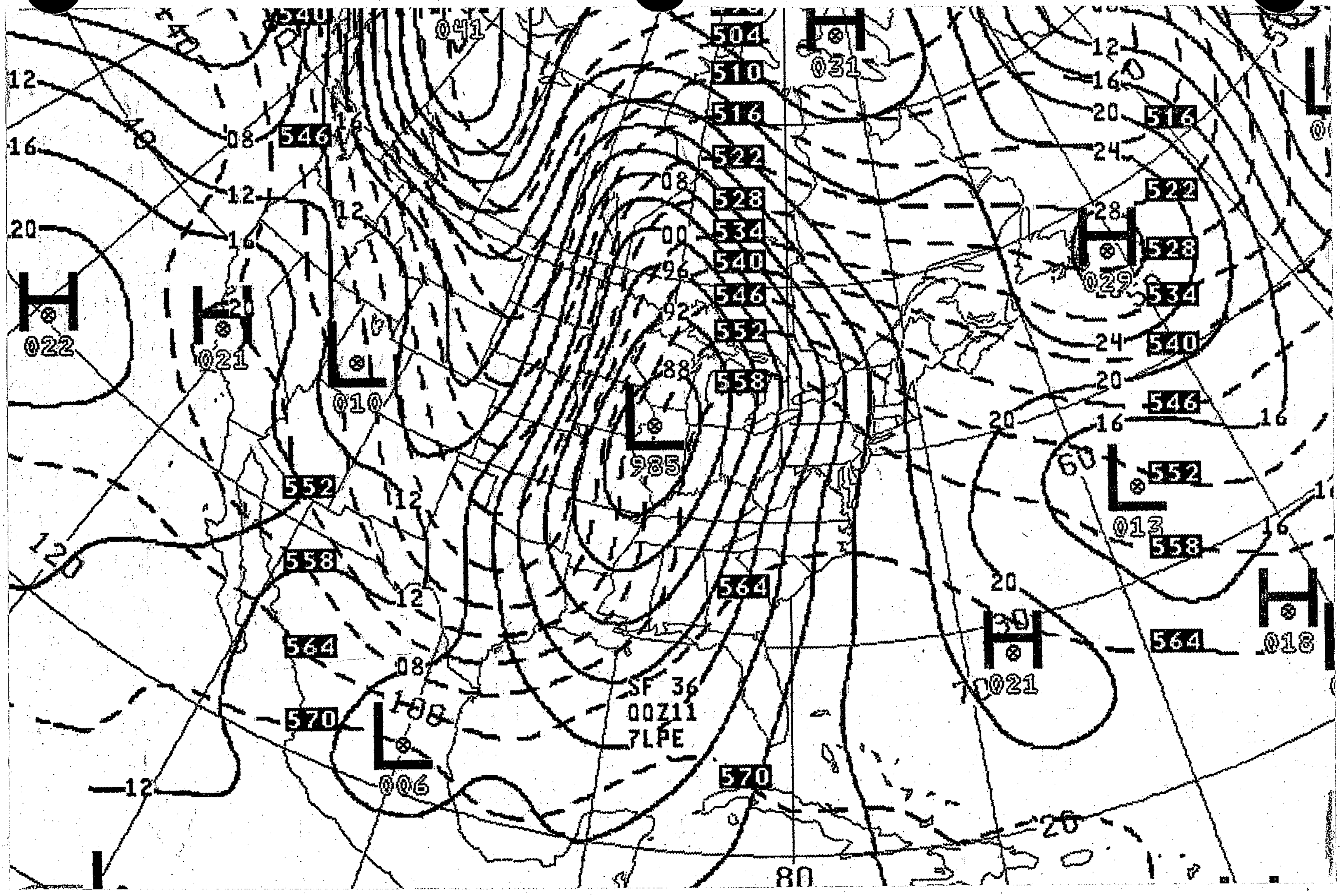
c

6LPE 48 hr forecast
valid 12Z 11 JAN 75

Fig. IV-8 - 7LPE 24, 36, 48 hr forecasts
Sea Level Pressure
500-1000 Thickness

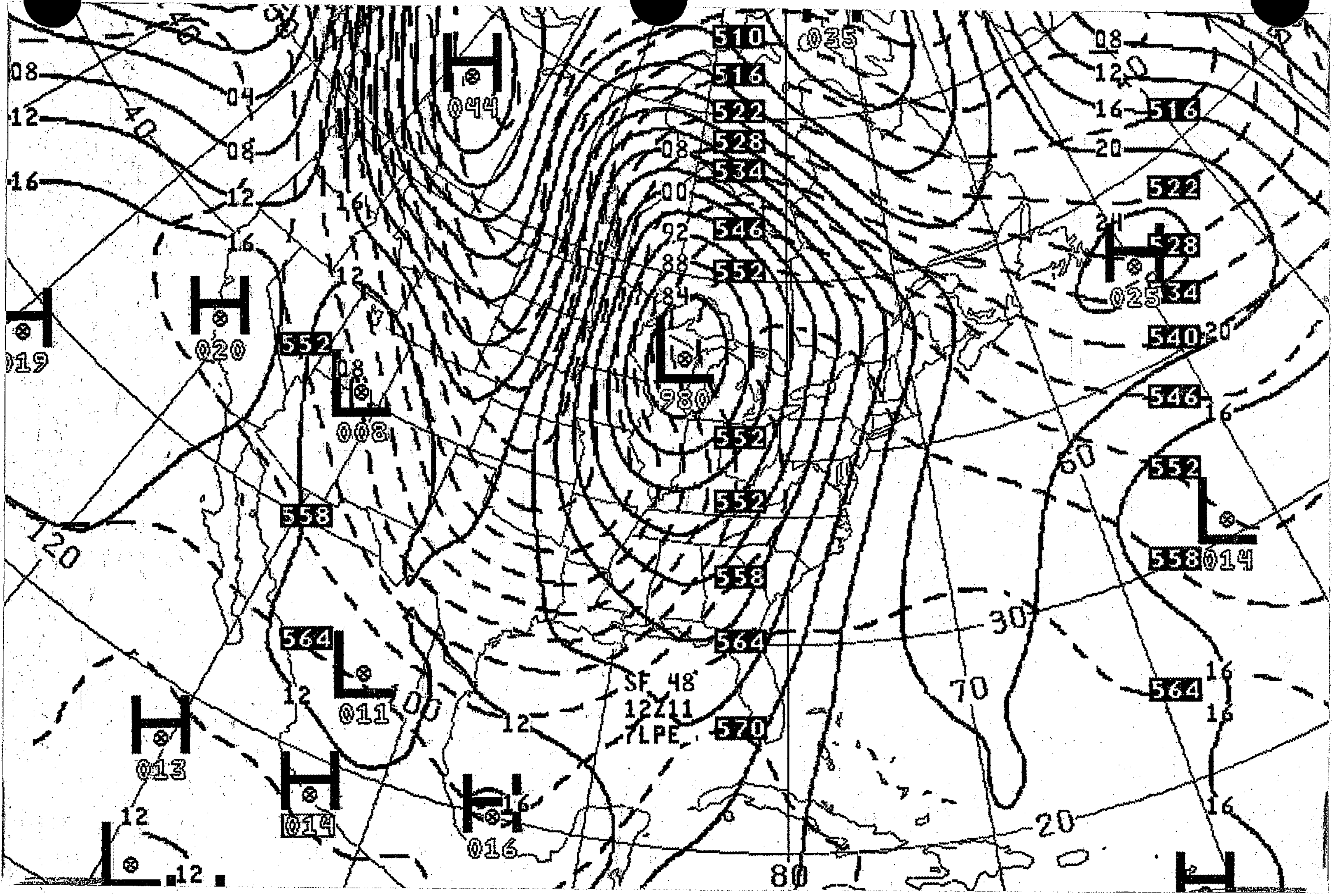


a
7LPE 24 hr forecast
valid 12Z 10 JAN 75



b

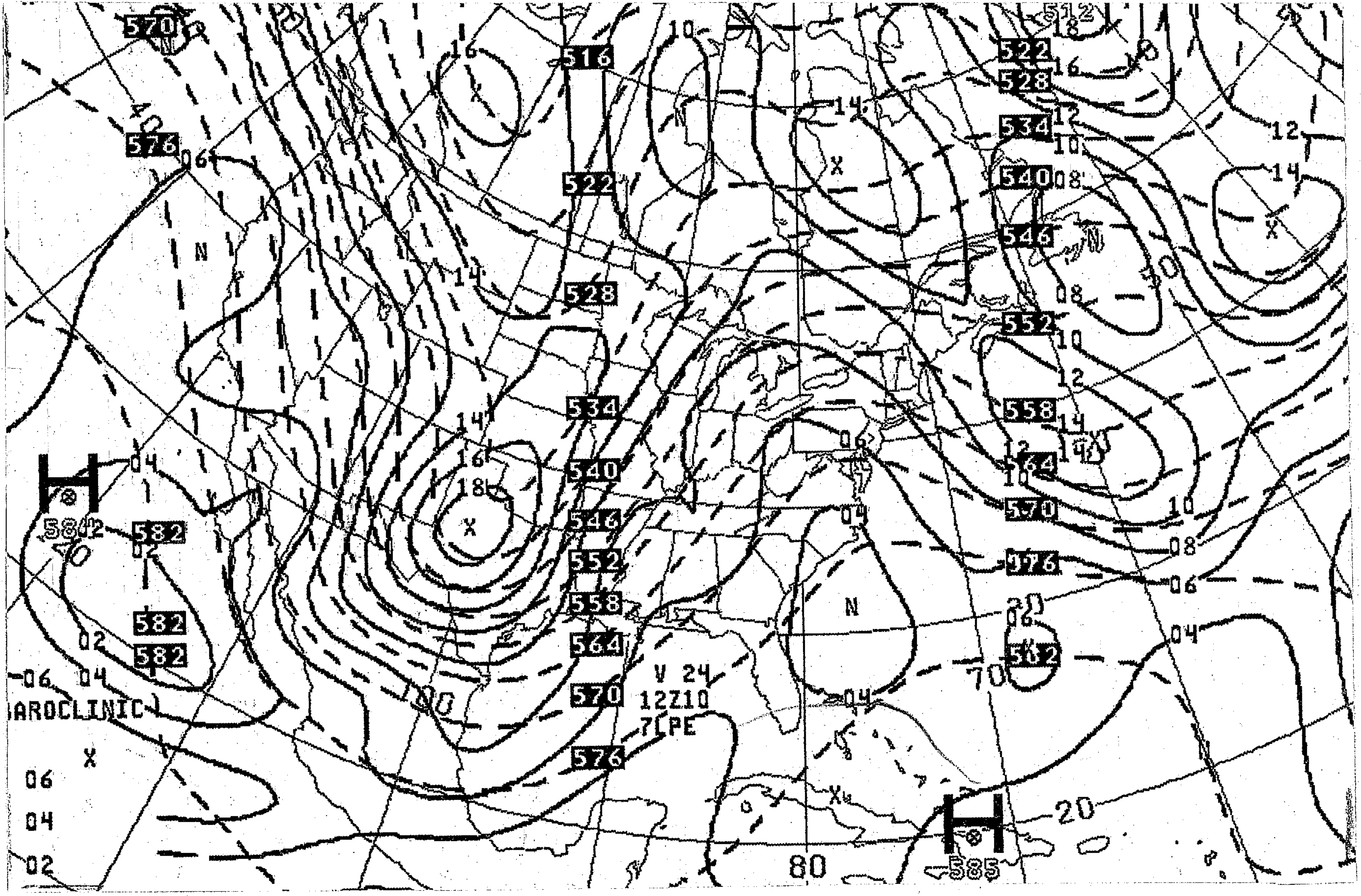
7LPE 36 hr forecast
valid 00Z 11 JAN 75



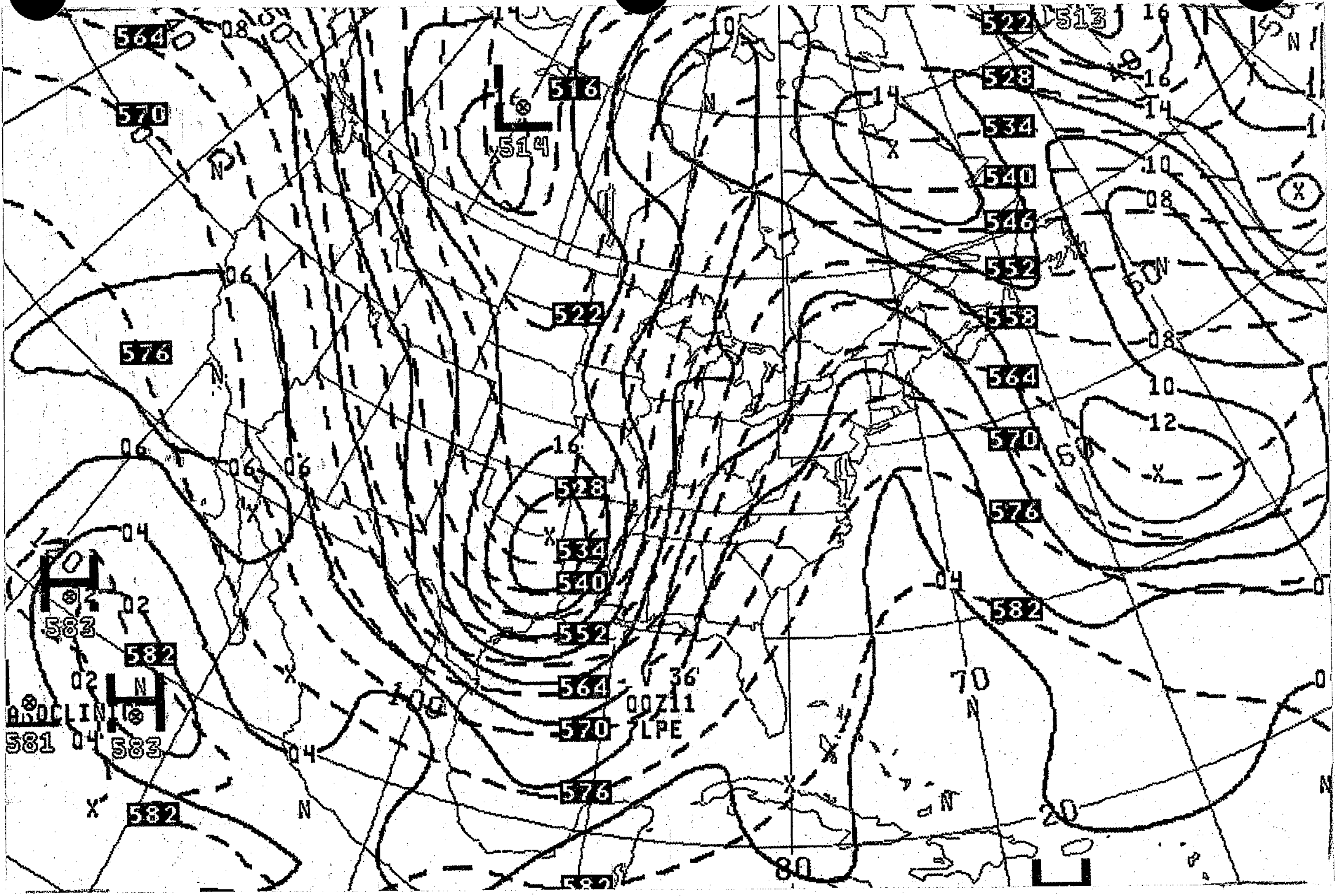
c

7LPE 48 hr forecast
valid 12Z 11 JAN 75

Fig. IV-9 - 7LPE 24, 36, 48 hr forecasts
500 mb height and vorticity

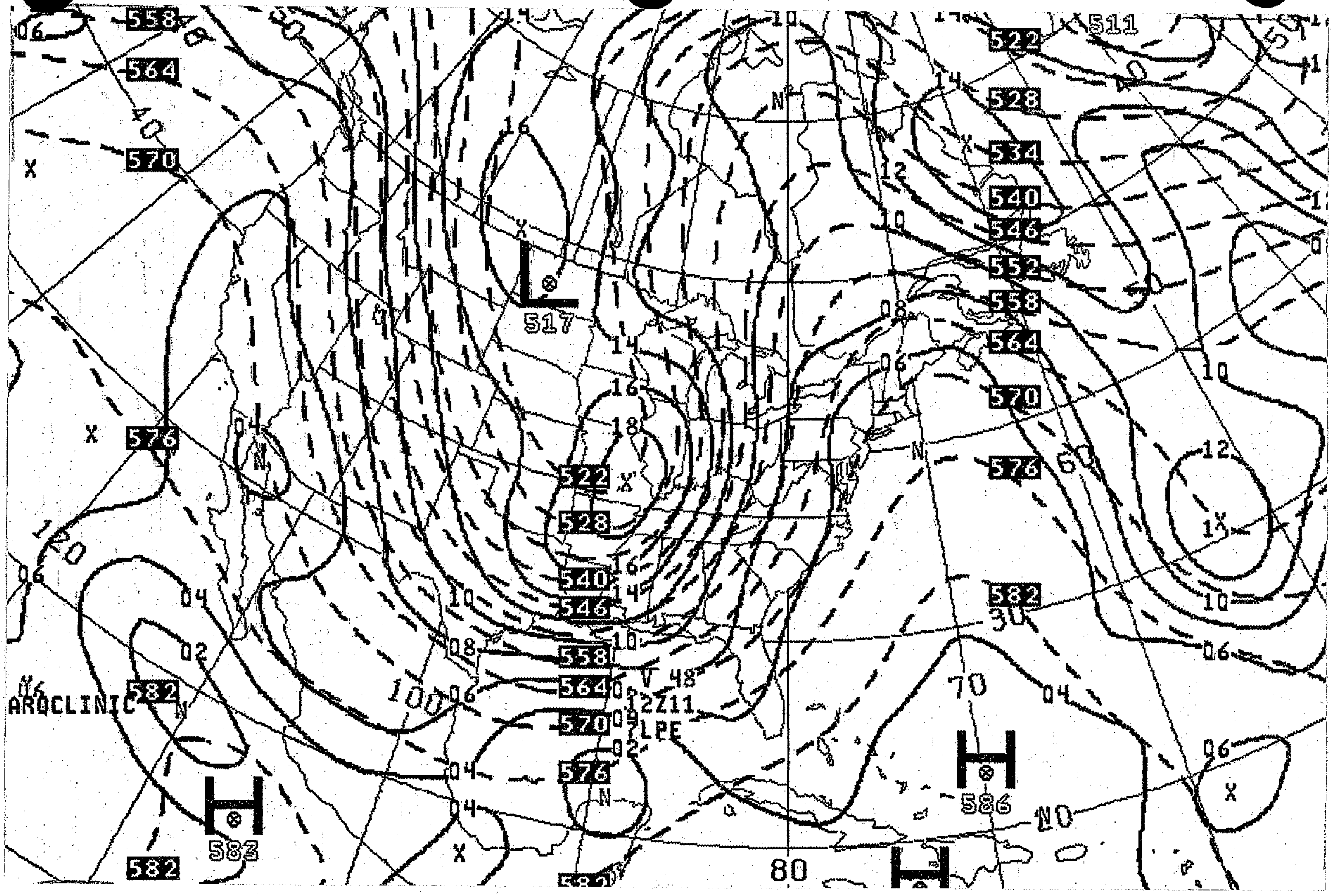


a
7LPE 24 hr forecast
valid 12Z 10 JAN 75



b

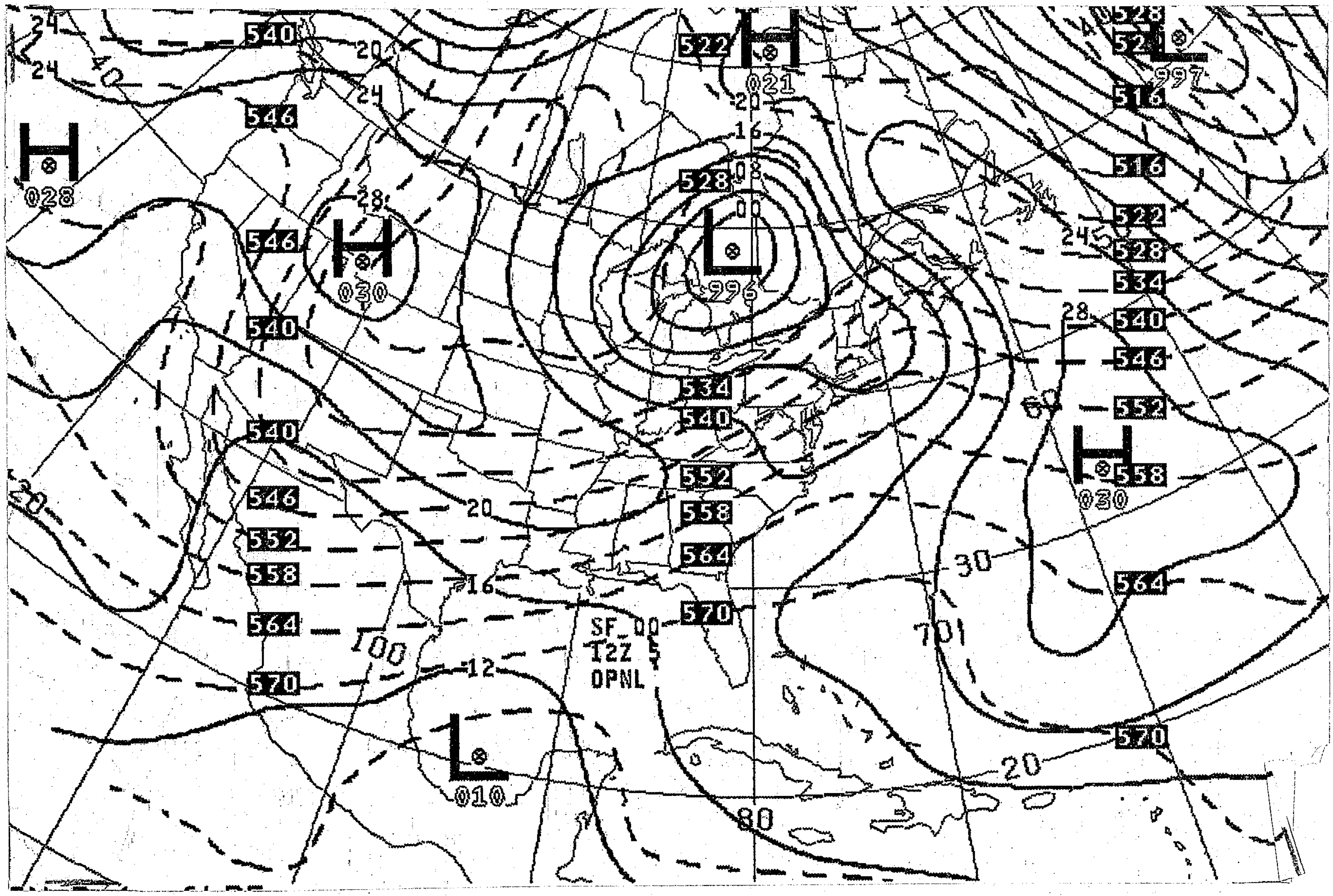
7LPE 36 hr forecast
valid 00Z 11 JAN 75



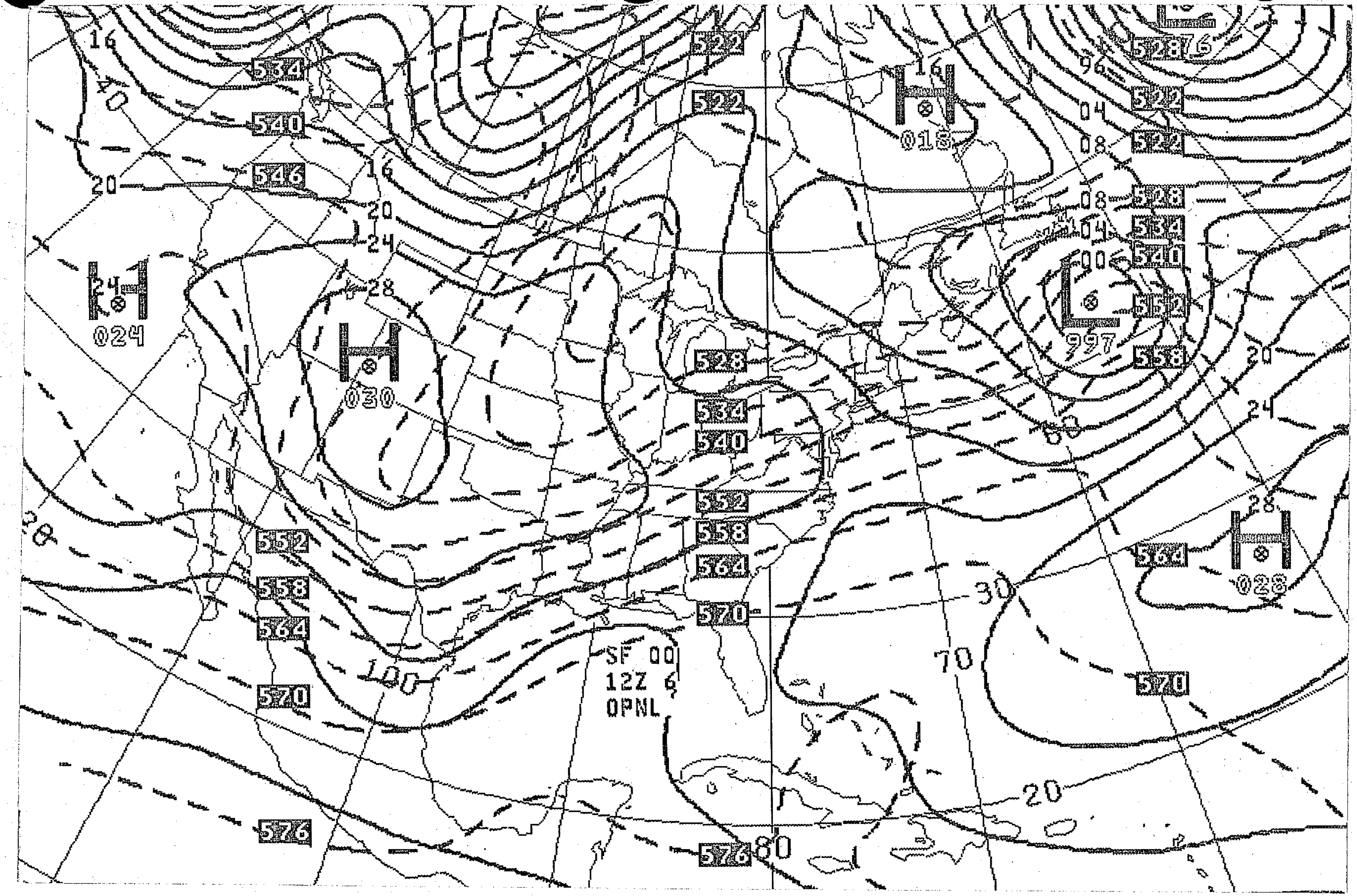
c

7LPE 48 hr forecast
valid 12Z 11 JAN 75

Fig. IV-10 - Analyses
Sea Level Pressure
500-1000 Thickness

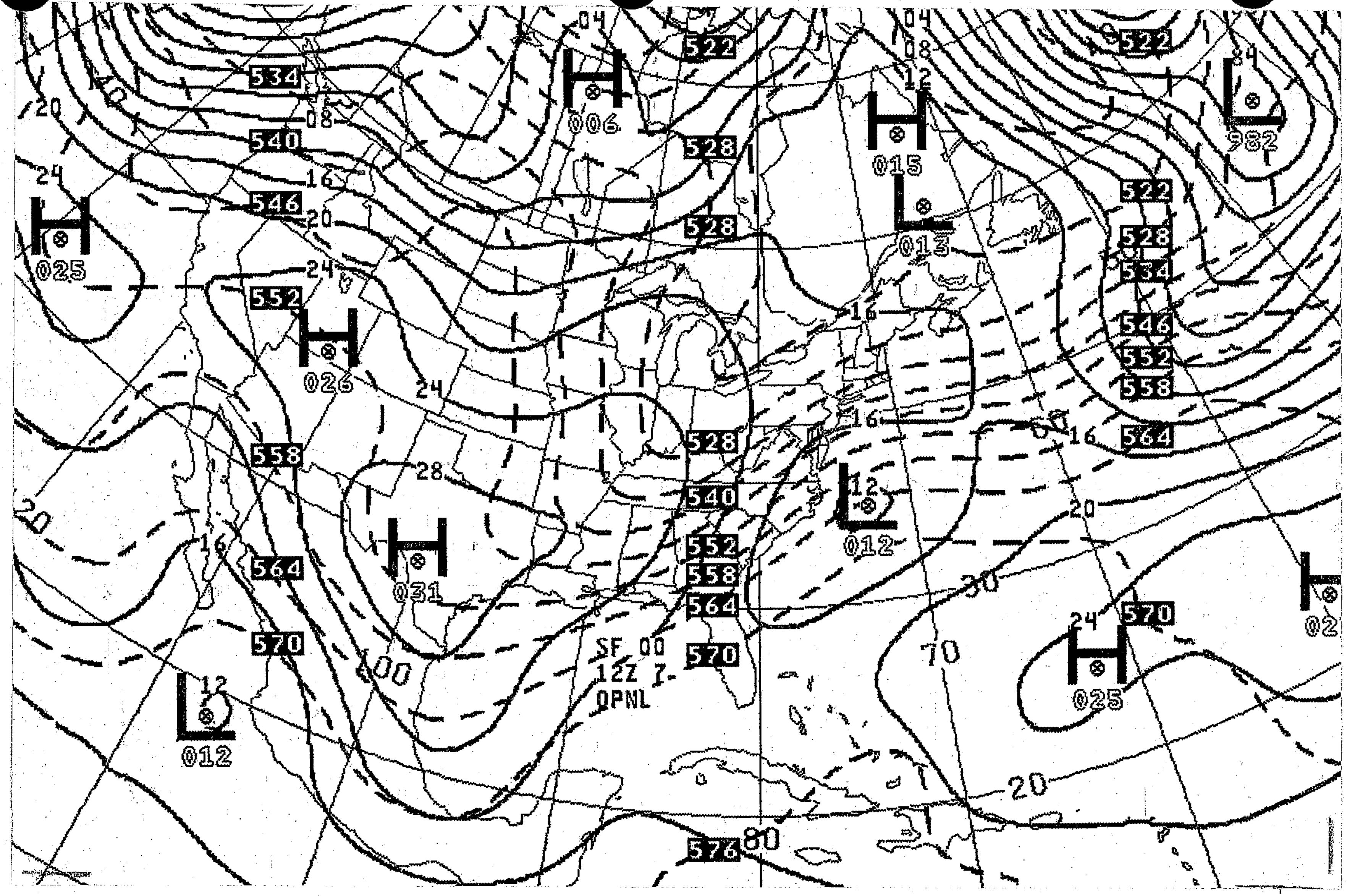


a
12Z 5 MAR 77


534
540
546
552
558
564
570
576
522
522
528
534
540
552
558
564
570
576
522
522
528
534
540
552
558
564
570
H
024
H
030
H
018
L
997
H
028
16
40
20
16
20
24
28
28
100
30
70
20
80
SF 00
12Z 6
OPNL

b

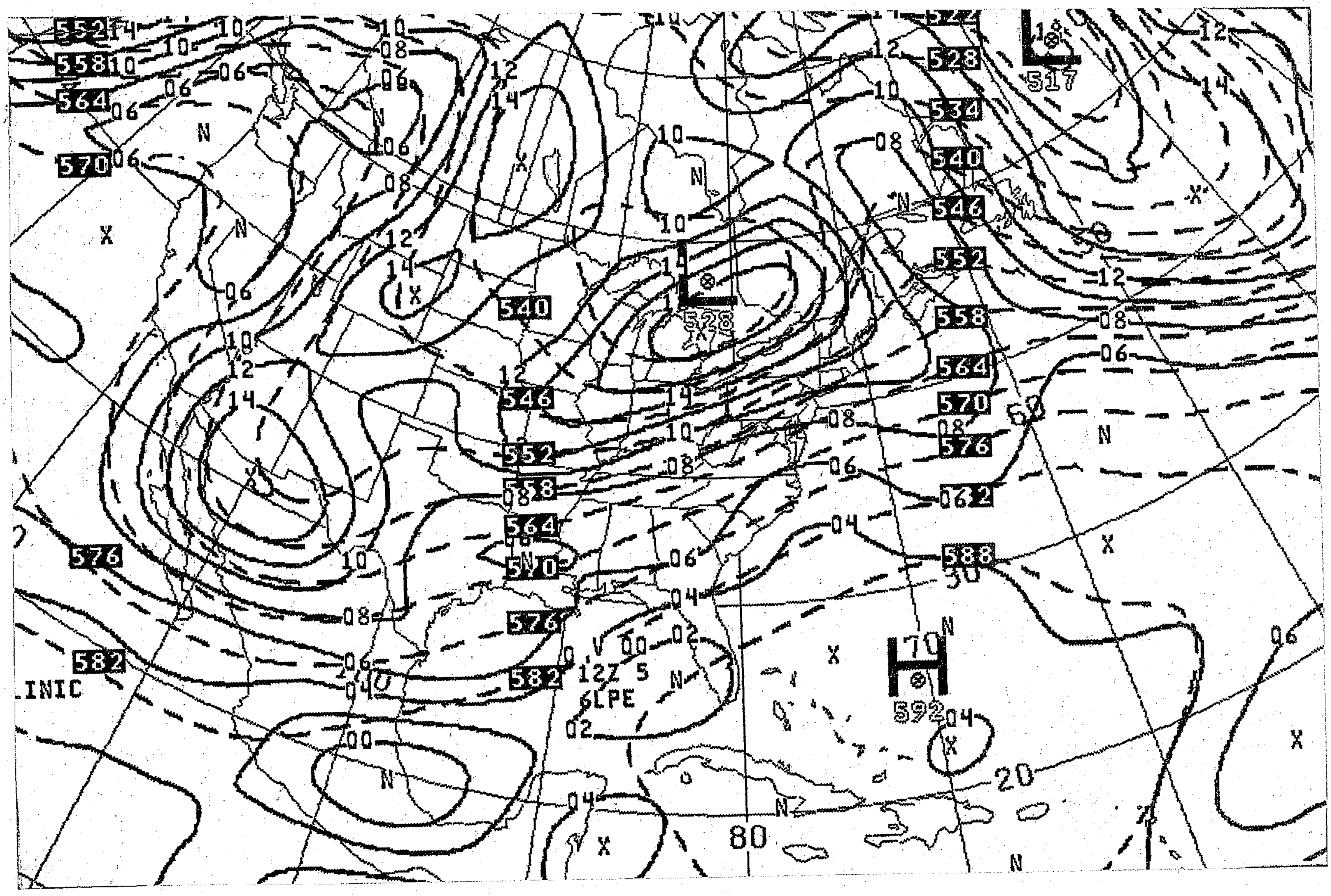
12Z 6 MAR 77


SF 00
12Z 7-
OPNL
H 006
H 015
L 013
L 982
H 025
H 026
H 031
L 012
L 012
H 025
522
528
534
540
546
552
558
564
570
576

c

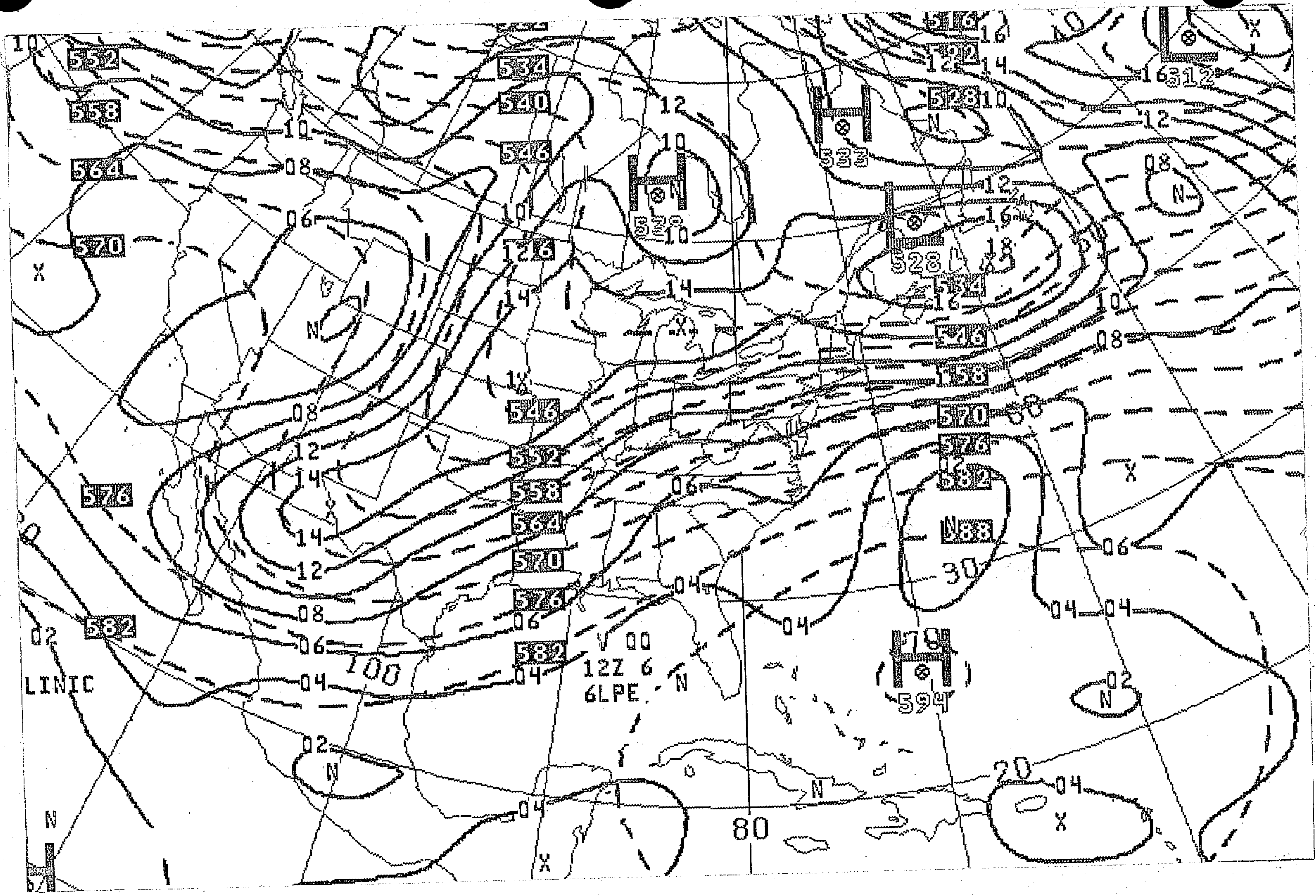
12Z 7 MAR 77

Fig. IV-11 - Analyses
500 mb height and vorticity



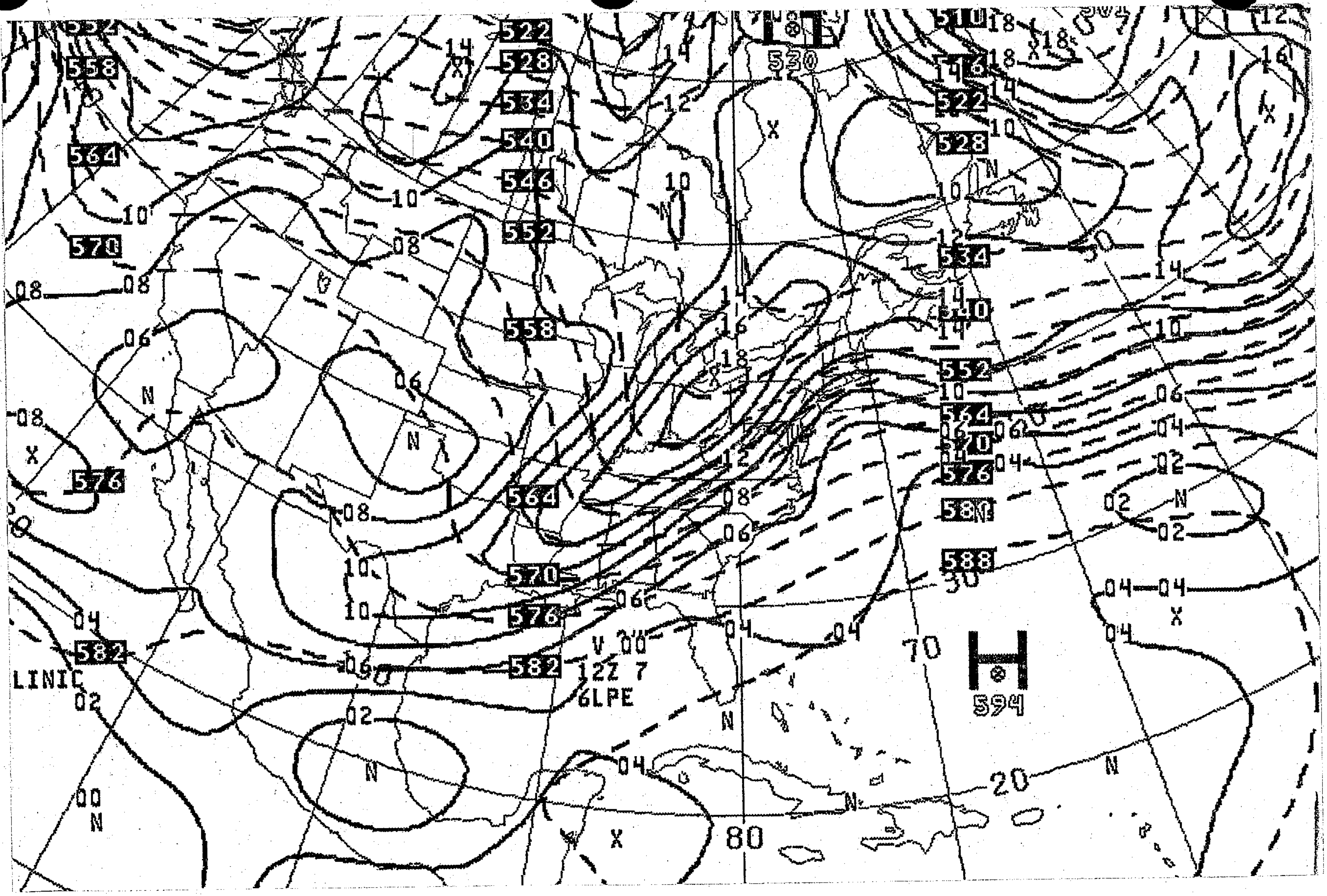
a

12Z 5 MAR 77


V 00
12Z 6
6LPE
LINIC
552
558
564
570
576
582
534
540
546
516
522
528
512
533
528
534
546
558
570
576
582
588
594
100
80
30
20
H
N
X
12
10
08
06
04
02
14
16
18

b

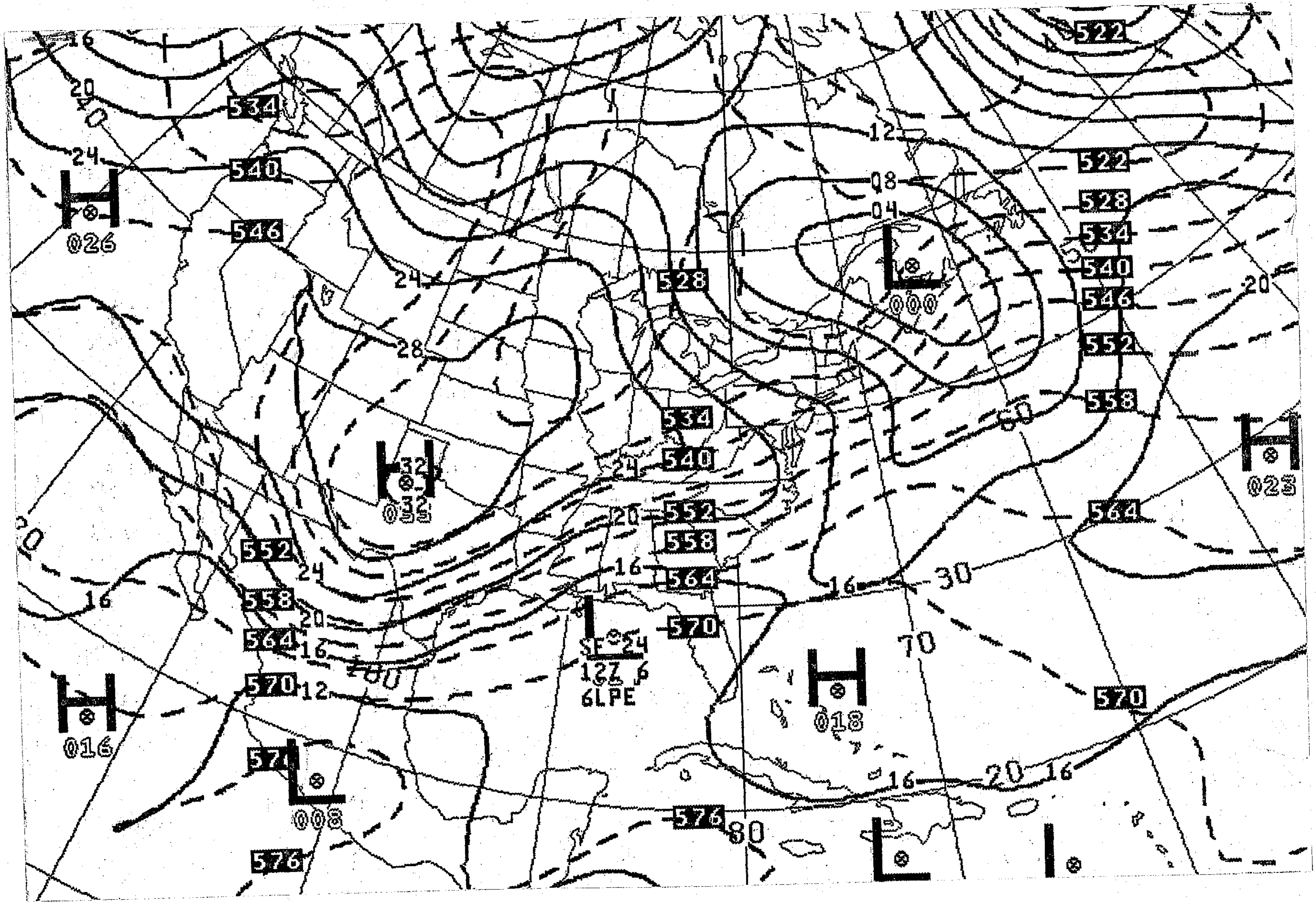
12Z 6 MAR 77


522
528
534
540
546
552
558
564
570
576
582
510
588
594
530
H
N
X
V 00
12Z 7
6LPE
LINIC
80
70
30
20

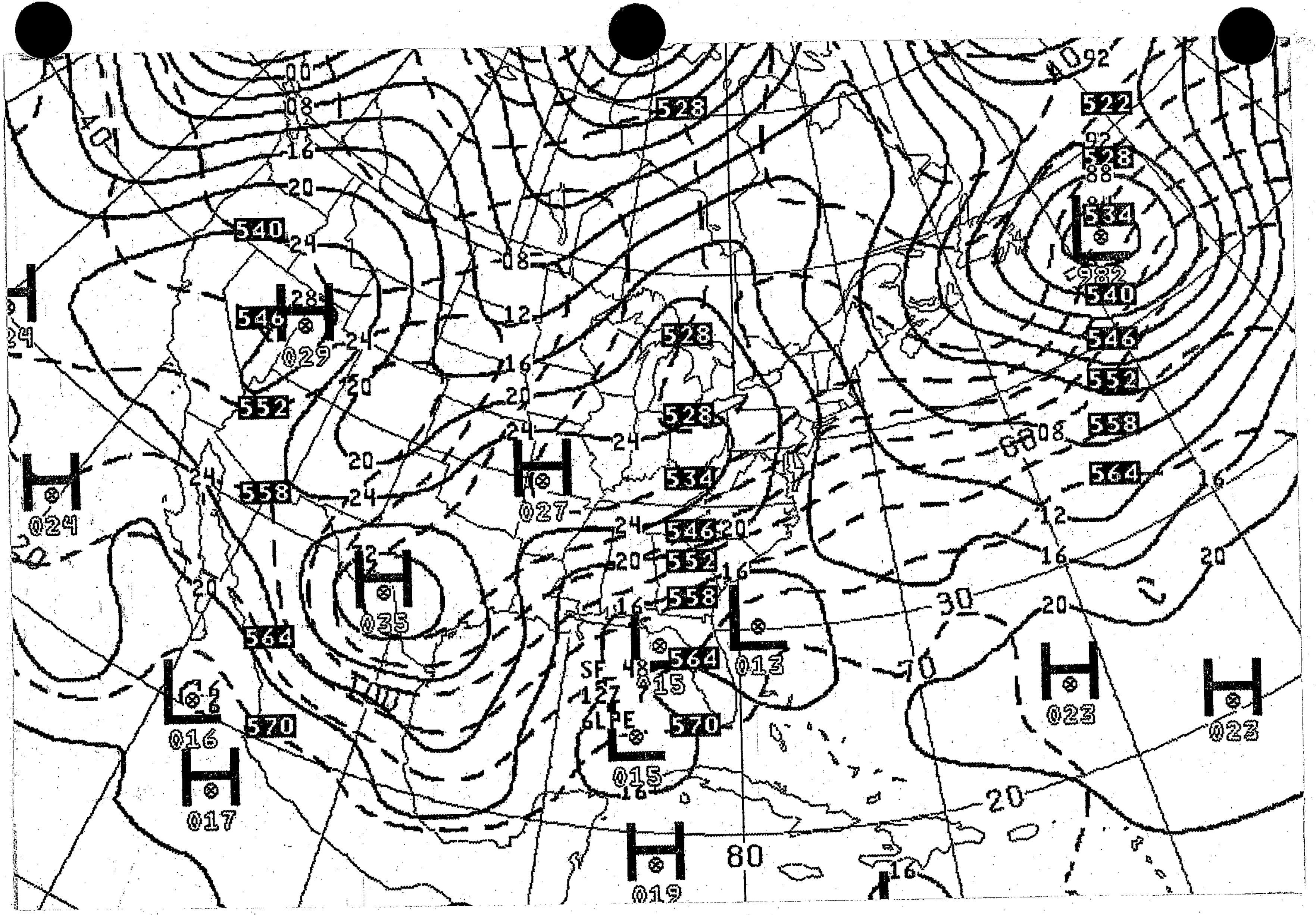
c

12Z 7 MAR 77

Fig. IV-12 - 6LPE 24, 48 hour forecasts
Sea Level Pressure
500-1000 Thickness



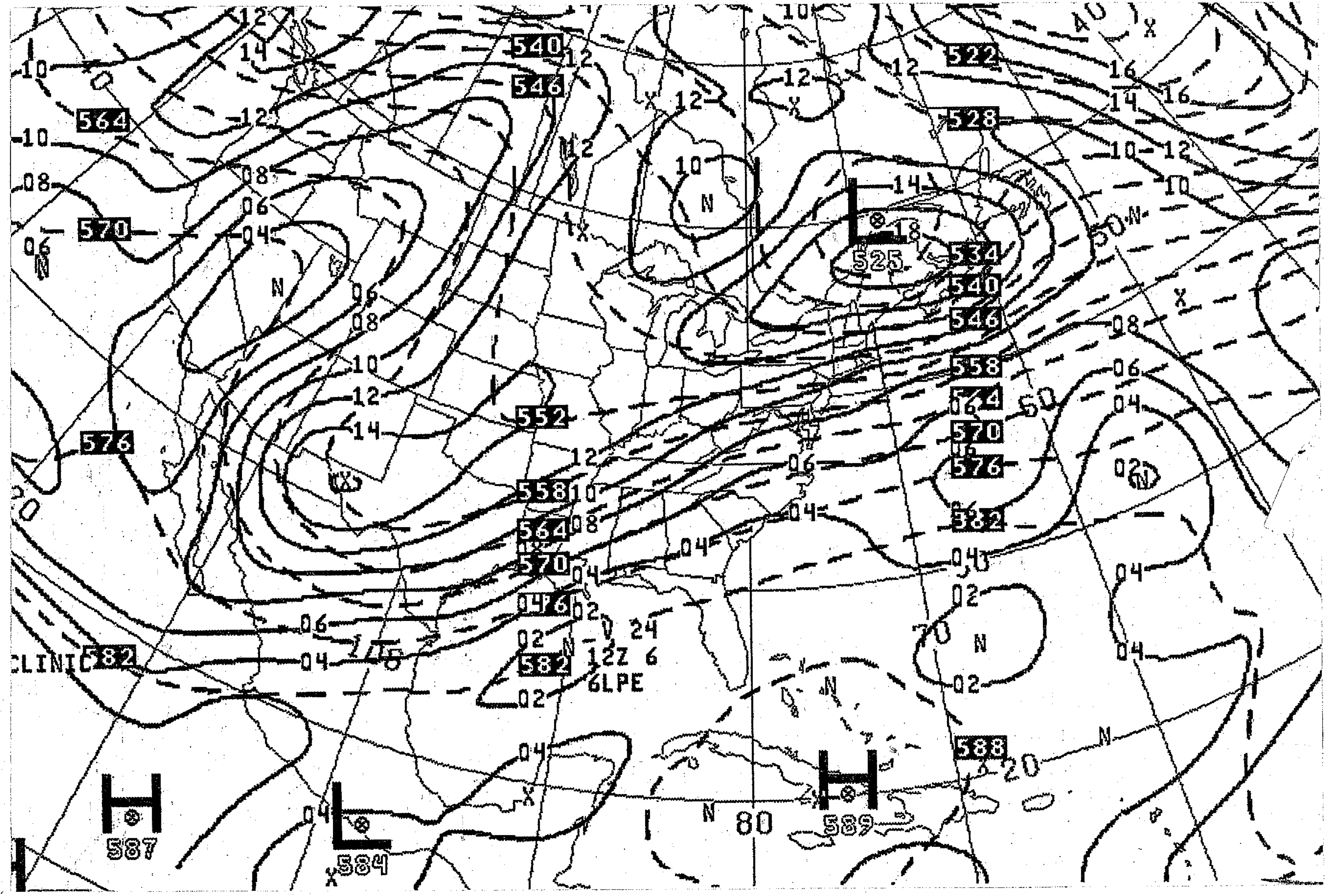
a
6LPE 24 hr forecast
valid 12Z 6 MAR 77



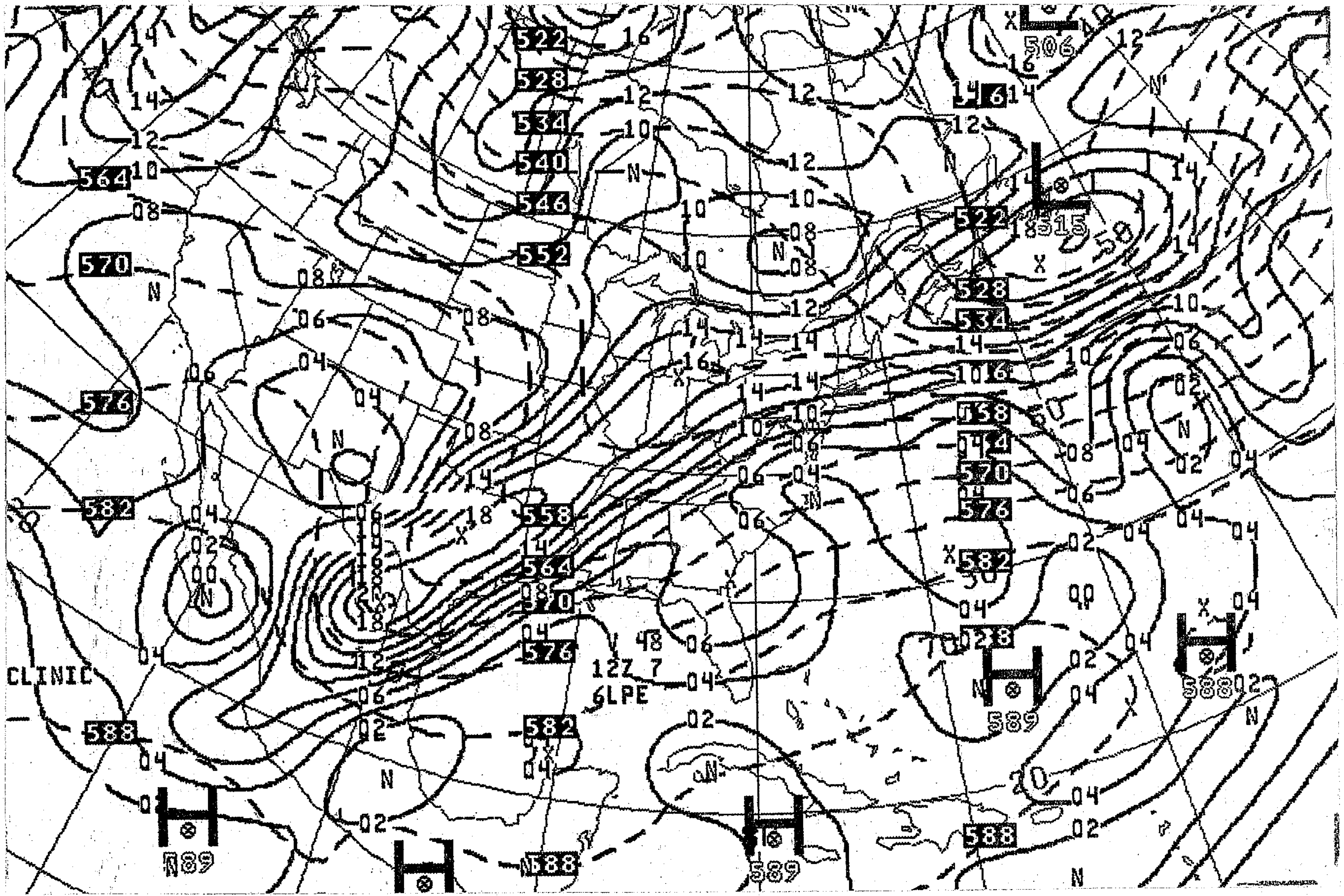
b

6LPE 48 hr forecast
valid 12Z 7 MAR 77

Fig. IV-13 - 6LPE 24, 48 hr forecasts
500 mb height and vorticity

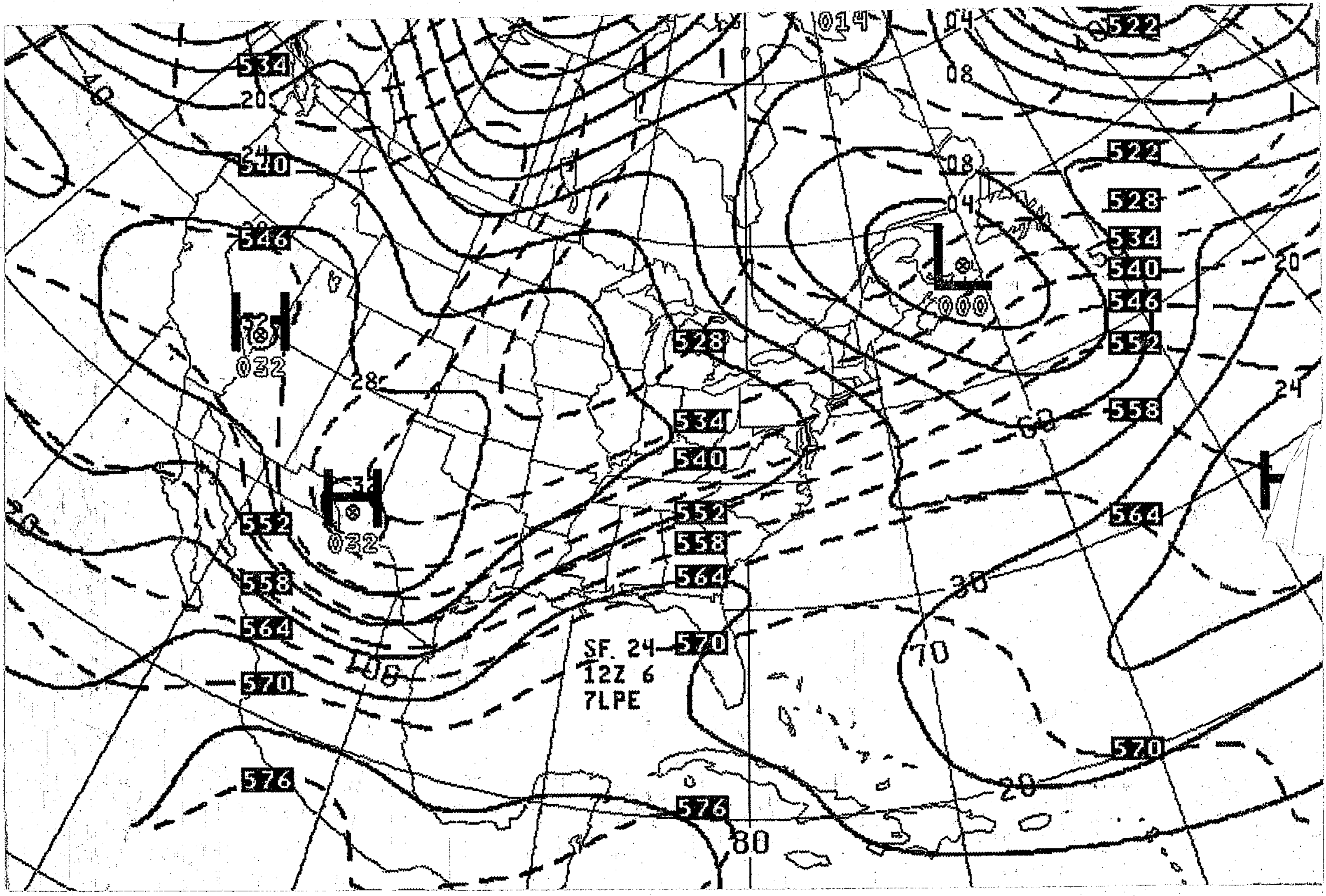


a
6LPE 24 hr forecast
valid 12Z 6 MAR 77

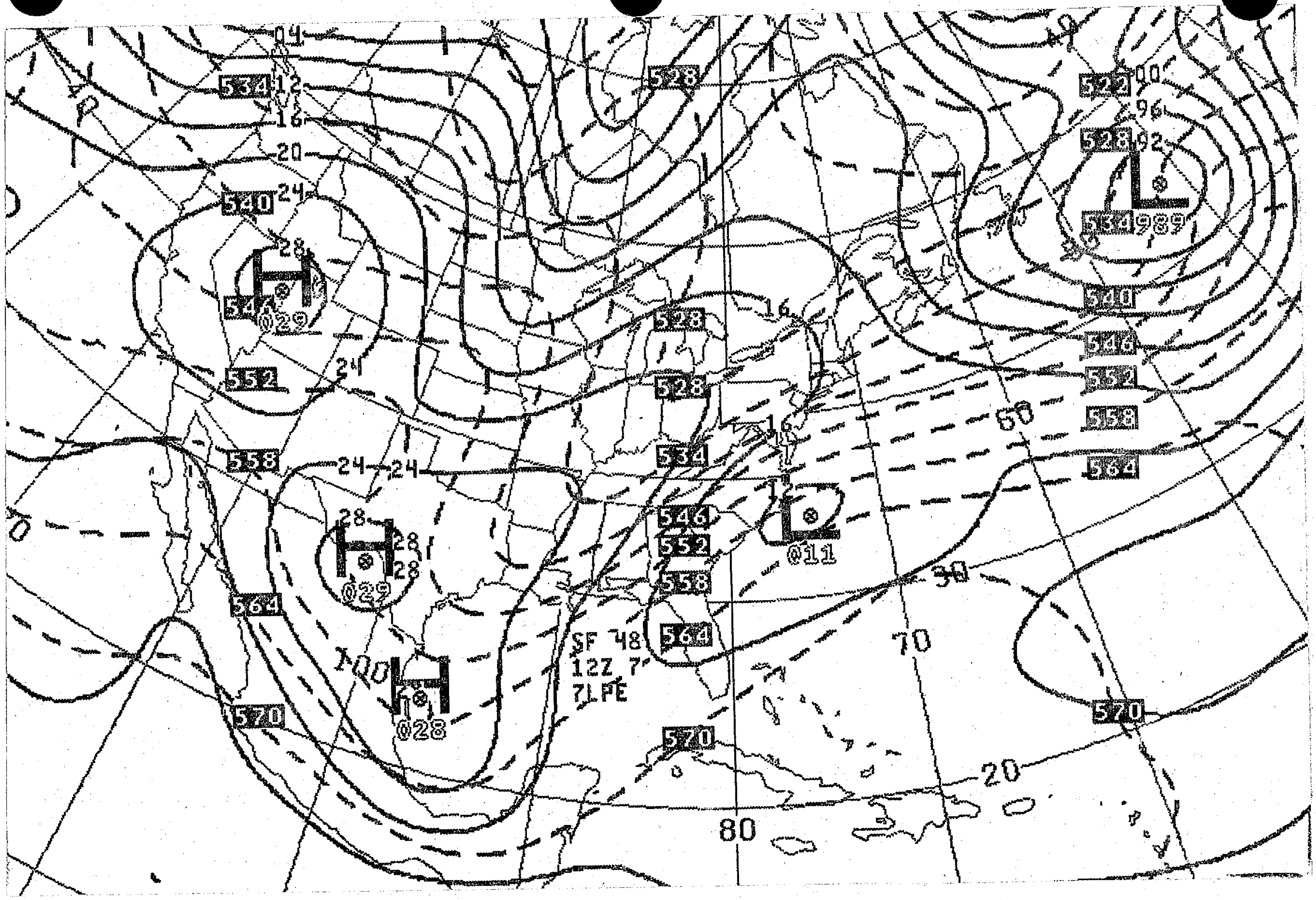


b
6LPE 48 hr forecast
valid 12Z 7 MAR 77

Fig. IV-14 - 7LPE 24, 48 hour forecasts
Sea Level Pressure
500-1000 Thickness



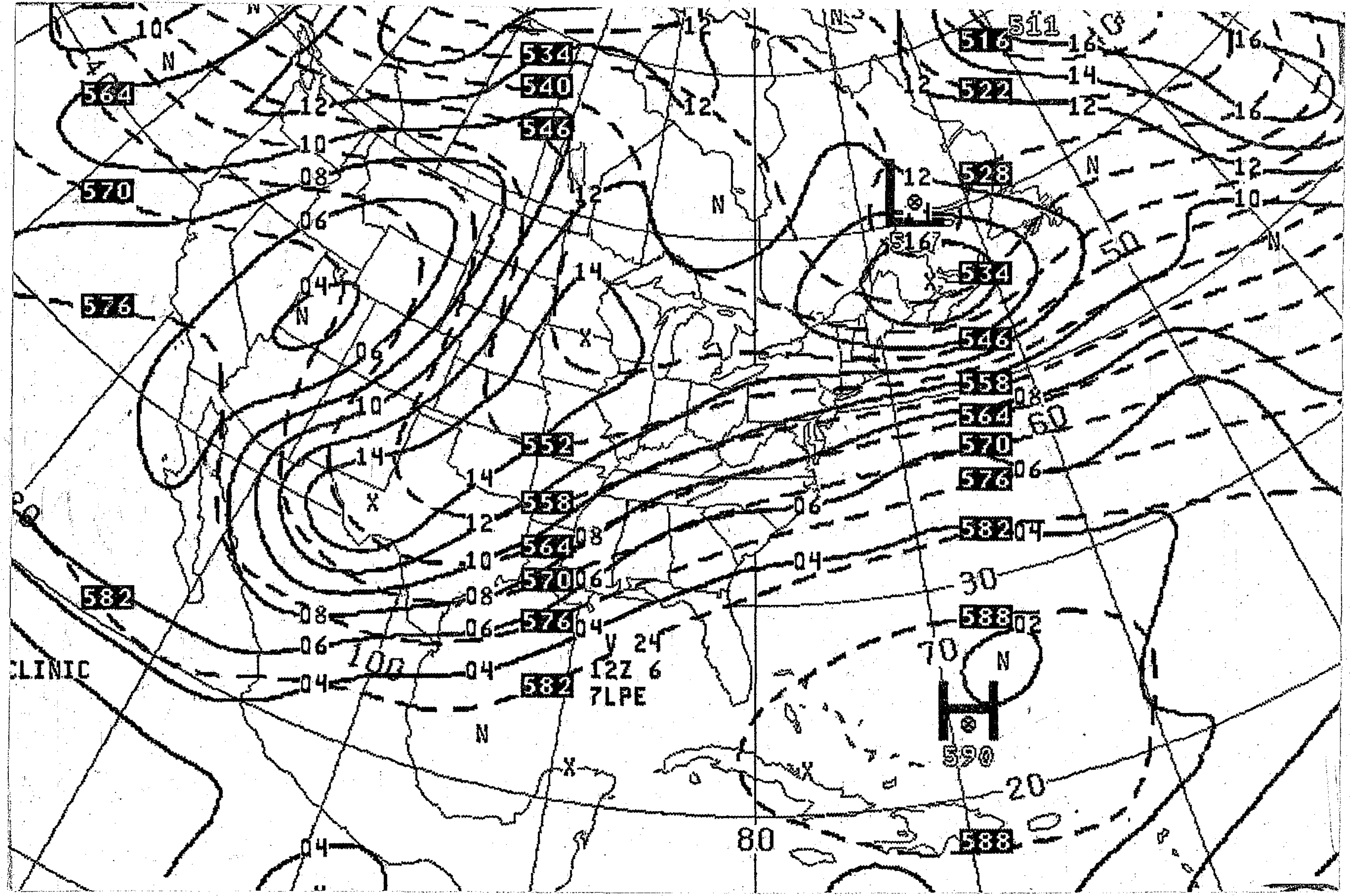
a
7LPE 24 hr forecast
valid 12Z 6 MAR 77



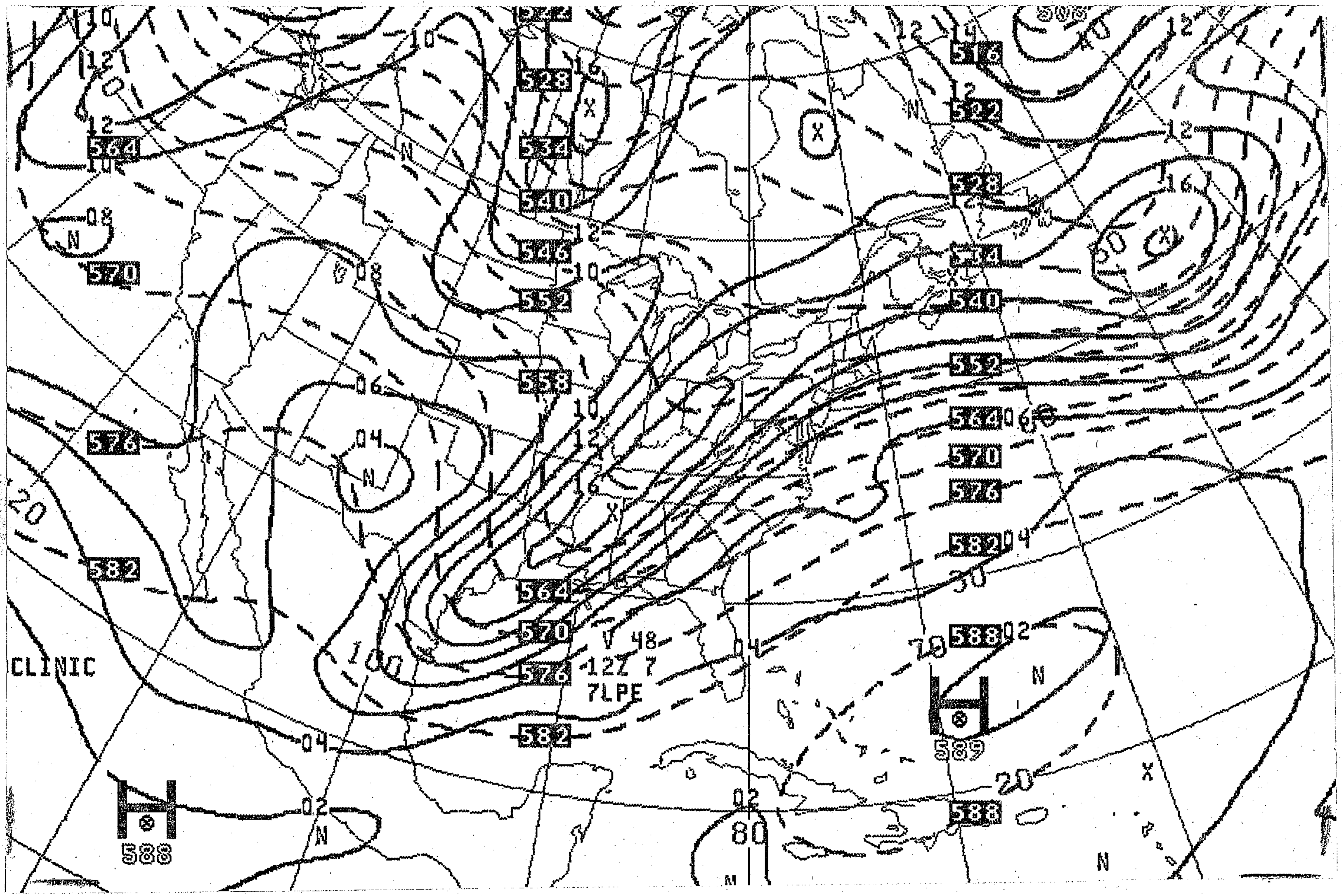
b

7LPE 48 hr forecast
valid 12Z 7 MAR 78

Fig. IV-15 - 7LPE 24, 48 hour forecasts
500 mb height and vorticity



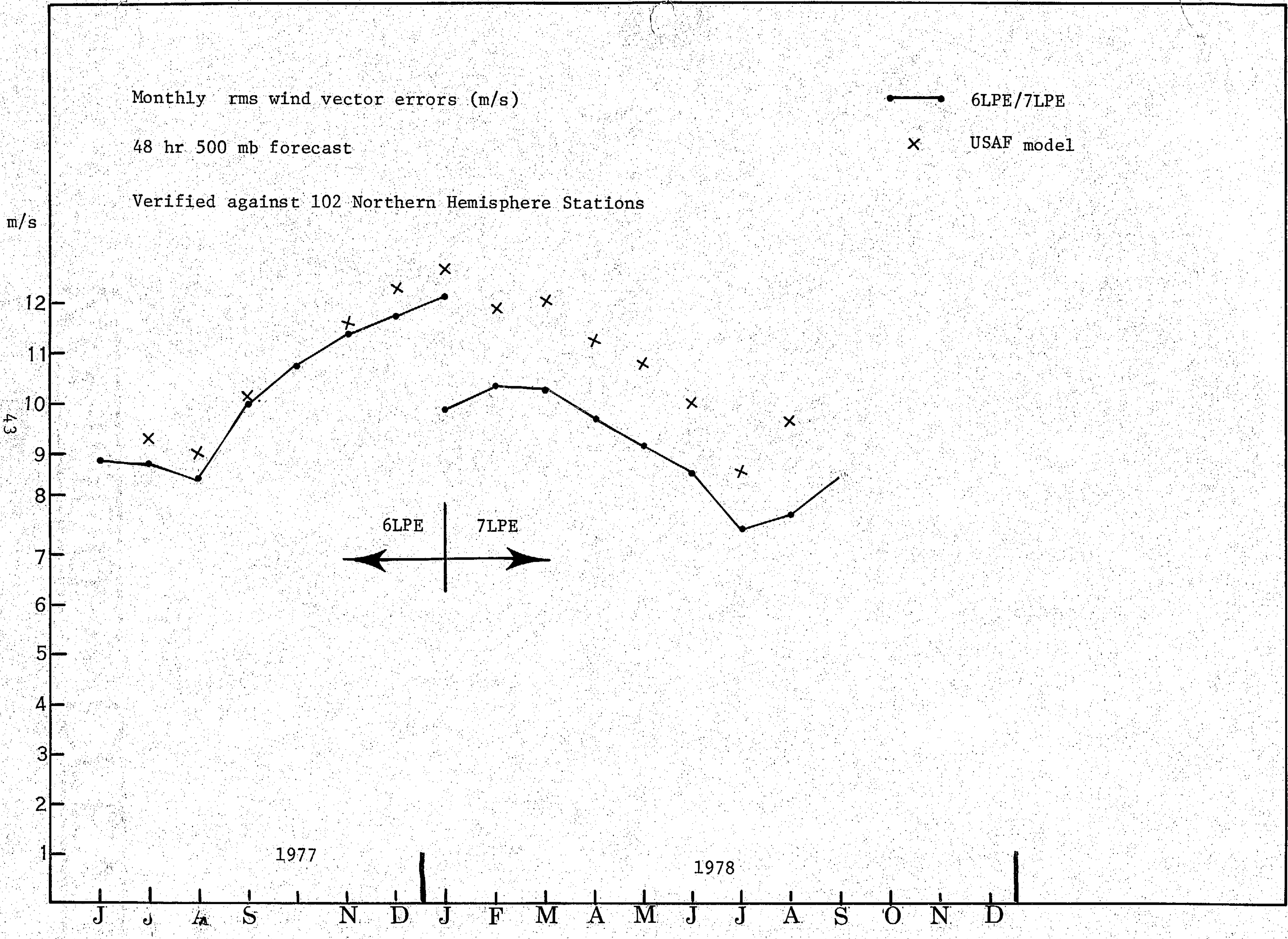
a
7LPE 24 hour forecast
valid 12Z 6 MAR 77



b

7LPE 48 hour forecast
valid 12Z 7 MAR 77

Fig. IV-16


Monthly rms wind vector errors (m/s)
48 hr 500 mb forecast
Verified against 102 Northern Hemisphere Stations
6LPE/7LPE
USAF model
m/s
12
11
10
9
8
7
6
5
4
3
2
1
6LPE
7LPE
1977
1978
J
J
A
S
N
D
J
F
M
A
M
J
J
A
S
O
N
D